JUL 3 1939

Mathematics Library

C-27

AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

EDITED BY

ABRAHAM COHEN
THE JOHNS HOPKINS UNIVERSITY

F. D. MURNAGHAN
THE JOHNS HOPKINS UNIVERSITY

T. H. HILDEBRANDT
UNIVERSITY OF MICHIGAN

J. F. RITT
COLUMBIA UNIVERSITY

R. L. WILDER
UNIVERSITY OF MICHIGAN

WITH THE COÖPERATION OF

G. C. EVANS
R. D. JAMES
GABRIEL SZEGÖ
AUREL WINTNER
LEO ZIPPIN
OYSTEIN ORE
H. P. ROBERTSON
M. H. STONE
T. Y. THOMAS
G. T. WHYBURN
E. T. BELL
H. B. CURRY
E. J. McSHANE
HANS RADEMACHER
OSCAR ZARISKI

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXI, Number 3
JULY, 1939

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

CONTENTS

	PAGE
Green's function and the problem of Plateau. By JESSE DOUGLAS,	545
The most general form of the problem of Plateau. By JESSE DOUGLAS,	590
Harmonic functions in domains with multiple boundary points. By JOHN W. GREEN,	609
Rigidity for spaces of class greater than one. By CARL B. ALLENDOERFER,	633
On a stretching process for surfaces. By T. RADÓ and P. REICHELDERFER,	645
On the problem of temperatures in a non-homogeneous bar with discontinuous initial temperatures. By R. V. CHURCHILL,	651
The number of circles covering a set. By RICHARD KERSHNER,	665
On Buffon's problem and its generalizations. By M. KAC, E. R. VAN KAMPEN and AUREL WINTNER,	672
Circular equidistributions and statistical independence. By M. KAC and E. R. VAN KAMPEN,	677
On oscillations of real sequences and of their transforms by square matrices. By RALPH PALMER AGNEW,	683
Groups with abelian norm quotient group. By REINHOLD BAER,	700
Condition that a finite group be multiply isomorphic with each of its irreducible representations. By LOUIS WEISNER,	709
Additive arithmetical functions and statistical independence. By PAUL ERDÖS and AUREL WINTNER,	713
On the smoothness of the asymptotic distribution of additive arithmetical functions. By PAUL ERDÖS,	722
On a characterization of the normal distribution. By M. KAC,	726
On the asymptotic distribution of a uniformly almost periodic function. By E. R. VAN KAMPEN,	729
Semi-locally connected sets. By G. T. WHYBURN,	733
Concerning local connectedness under the inverse of certain continuous transformations. By W. T. PUCKETT, JR.,	750
Some characterizations of interior transformations. By A. D. WALLACE,	757
A metric connected with property S. By J. L. KELLEY,	764
Über einen Tauberschen Satz im Dreikörperproblem. By J. KARAMATA,	769
Infinite number fields with Noether ideal theories. By SAUNDERS MAC LANE and O. F. G. SCHILLING,	771
Ring homomorphisms which are also lattice homomorphisms. By MORGAN WARD,	783
Note on the general rational solution of the equation $ax^2 - by^2 = z^3$. By MORGAN WARD,	788
On extensors and a Euclidean basis for higher order spaces. By H. V. CRAIG,	791

THE AMERICAN JOURNAL OF MATHEMATICS will appear four times yearly.

The subscription price of the JOURNAL for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the JOURNAL remain on sale.

Papers intended for publication in the JOURNAL may be sent to any of the Editors.

Editorial communications may be sent to Dr. A. COHEN at The Johns Hopkins University.

Subscriptions to the JOURNAL and all business communications should be sent to THE JOHNS HOPKINS PRESS, BALTIMORE, MARYLAND, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

GREEN'S FUNCTION AND THE PROBLEM OF PLATEAU.*

By JESSE DOUGLAS.

1. INTRODUCTION.

1. This paper presents the details of a particularly simple and general solution of the general topological form of the Plateau problem, based on the following three formulas involving the Green's function of an arbitrary Riemann surface R. The formulas are:

$$(1.1)\qquad A(g,R)=\frac{1}{8\pi}\int_C\int_C[g(P_1)-g(P_2)]^2\frac{\partial^2 G(P_1,P_2)}{\partial n_1\partial n_2}\,ds_1\,ds_2.$$

$$(1.2)\qquad \frac{\partial H(Q)}{\partial\xi}\cdot\frac{\partial H(Q)}{\partial\eta}$$
$$=-\frac{1}{16\pi^2}\int_C\int_C[g(P_1)-g(P_2)]^2\frac{\partial^2}{\partial n_1\partial n_2}\left\{\frac{\partial G(P_1,Q)}{\partial\xi}\,\frac{\partial G(P_2,Q)}{\partial\eta}+\frac{\partial G(P_1,Q)}{\partial\eta}\,\frac{\partial G(P_2,Q)}{\partial\xi}\right\}ds_1\,ds_2.$$

$$(1.3)\qquad \delta G(P_1,P_2)=\frac{\partial G(P_1,Q)}{\partial\xi}\,\frac{\partial G(P_2,Q)}{\partial\eta}+\frac{\partial G(P_1,Q)}{\partial\eta}\,\frac{\partial G(P_2,Q)}{\partial\xi}.$$

Here Q denotes an arbitrary point of the Riemann surface R, and ξ, η are any two perpendicular directions through Q. *The letters g, H represent vectors with any number n of components, and will continue to do so throughout this paper.* The dot denotes the scalar product of two vectors, and the exponent two the scalar product of a vector by itself (sum of squares of components). δ is the characteristic symbol of the calculus of variations. The other notation and general background will be explained in the sequel.

An account of the essential features of the present paper, including the statement of these three fundamental formulas and a quotation of the main theorem—the Variational Theorem of § 5. 4, has already appeared in the Proceedings of the National Academy of Sciences [11].[1]

The most important part of the paper is § 5, which proves this Variational Theorem, involving the formula (1. 3).

* Received August 15, 1938.

[1] Numbers in square brackets are citations from the list of references at the end of this paper.

2. The author has formulated and solved [2] the problem of Plateau in the following topologically general form: *given*

(1) *any finite number k of contours,*[3] Γ, *in the form of Jordan curves in n-dimensional euclidean space, with assigned form, position, and sense of description*;

(2) *any prescribed genus h or topological characteristic* [4] r;

(3) *either character of orientability, i. e., one-sided or two-sided*;

to determine a minimal surface M belonging to these data.

The existence of the solution depends on certain simple sufficient conditions, first given by the author,[5] which, in their most concrete form, express that the lower bound of the areas of all surfaces corresponding properly to the data (1), (2), (3) is *less than* the similar lower bound for separated surfaces with total characteristic r' of the component parts $\leqq r$, or for non-separated surfaces of characteristic $< r$.[6]

Two forms of solution were given. One depended on the area functional and utilized the conformal mapping of polyhedral surfaces.[7] The other, playing the central rôle, and really typical of the author's method, was based on the "vector Dirichlet functional"

$$(1.4)\quad D(\mathfrak{x}) = \tfrac{1}{2}\int\int (\mathfrak{x}_u^2 + \mathfrak{x}_v^2)\, du\, dv = \tfrac{1}{2}\int\int \sum_{i=1}^{n}\left[\left(\frac{\partial x_i}{\partial u}\right)^2 + \left(\frac{\partial x_i}{\partial v}\right)^2\right] du\, dv.$$

and did not employ area or the theory of conformal mapping; in fact, for $n = 2$, this method gave as corollary the classic theorems [8] concerning the conformal mapping of plane regions.

---

[2] In the series of papers listed at the end of the present one.

[3] The symbol Γ will denote throughout the *ensemble* of given contours.

[4] This is the maximum number of circuits, no linear combination of which separates the surface. r may also be interpreted as the excess of the connectivity over the number of boundaries; $c = r + k$. For two-sided surfaces of genus h, $r = 2h$. For one-sided surfaces, r may have any integer value, odd or even, $\geqq 1$. Examples: Möbius strip with h handles, $r = 2h + 1$; Klein surface with h handles, $r = 2h + 2$. See Hilbert and Cohn-Vossen, *Anschauliche Geometrie*, 1932.

[5] In [7], Theorems I and II.

[6] See § 8.7. (Refer to footnote 10.)

[7] [9], § 19. Also, for the one-contour case, "The mapping theorem of Koebe and the problem of Plateau," *Journal of Mathematics and Physics*, v. 10 (1931), pp. 106-130; as well as [2], § 26.

[8] Of Riemann, Schottky, and Osgood-Carathéodory.

Since, for given boundary values, a harmonic function gives the least value to the Dirichlet integral, then, in considering the problem

$$D(\mathfrak{x}) = \min. \tag{1.5}$$

for all vectors [9] $\mathfrak{x}(u, v)$ that represent a surface of a given topological type bounded by Γ, we may, without affecting the solution, confine ourselves to harmonic vectors $\mathfrak{x} = H(u, v)$:

$$D(H) = \min. \tag{1.6}$$

The (u, v) domain on which the vector-functions are defined is always a Riemann surface R of the topological form prescribed for the required minimal surface M: k boundaries C without common points, and characteristic r. According to the Dirichlet principle, the harmonic function $H(u, v)$ on R is completely determined by its boundary values g on C. Hence $D(H)$ is a functional of g and R, for which we have introduced the notation

$$A(g, R) = D(H). \tag{1.7}$$

The explicit formula for $A(g, R)$, directly in terms of its arguments, is given by (1.1), which will be derived in detail later (§ 4.2).[10] One advantage of the solution based upon it is its *generality*; the method proceeds *entirely regardless of the particular topological form of R, i. e.,* of the number k of given contours and the characteristic r of the required minimal surface M. Indeed, as is pointed out in § 6, *the case is included where R or M is a perfectly general riemannian manifold, with an infinite number of boundaries and an infinite type of connectivity.*[10a]

We need hardly call attention again to the fact that also the number n of coördinates of the vector g is altogether indifferent; [11] if we wish, g may even have a denumerable infinity of coördinates, i. e., the Plateau problem may be considered in Hilbert space,[12] without the slightest additional difficulty.

3. Exact formulation. By a *minimal surface* we shall understand a surface M in n-dimensional euclidean space $\mathfrak{x}$, defined as a *harmonic* and

---

[9] Piece-wise continuously differentiable.

[10] This is our mode of designation of section and article, the former preceding, the latter following the dot.

[10a] See the author's paper immediately following the present one.

[11] This has been one of the typical features of the author's work from the start; and, among other advantages, is what permits us, by putting $n = 2$, to include the conformal mapping of plane regions.

[12] The space of all infinite sequences $(x_1, x_2, \cdots, x_m, \cdots)$ where $x_1^2 + x_2^2 + \cdots + x_m^2 + \cdots$ is convergent.

conformal image of any Riemann surface R. The formulas which express this definition are

$$\mathfrak{x} = H(u, v), \tag{1.8}$$

$$E = G, \quad F = 0, \tag{1.9}$$

where E, F, G are the fundamental quantities of the surface H:

$$E = \left(\frac{\partial H}{\partial u}\right)^2, \qquad F = \frac{\partial H}{\partial u} \cdot \frac{\partial H}{\partial v}, \qquad G = \left(\frac{\partial H}{\partial v}\right)^2. \tag{1.10}$$

These are evidently equivalent to the formulas given by Weierstrass for $n = 3$ and a simply-connected minimal surface, namely:

$$x_i = \Re F_i(w), \qquad (i = 1, 2, \cdots, n) \tag{1.11}$$

$$\sum_{i=1}^{n} F'_i{}^2(w) = 0, \tag{1.12}$$

for, after identifying (1.11) with (1.8), we have

$$F'_i(w) = \frac{\partial H_i}{\partial u} - i \frac{\partial H_i}{\partial v}. \tag{1.13}$$

We may also express the conformal relation between R and M by requiring only that $H(u, v)$ transform all *right angles* on R into *right angles* on M. Indeed, as is well-known from the general theory of maps,[13] it is even sufficient to require that two distinct right angles at each point of R remain right angles in the transformation to M.

If, then, Q denote an arbitrary point of R and ξ, η any two perpendicular directions through Q, the conditions (1.9) may be replaced by

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0. \tag{1.14}$$

In fact, $F = 0$ expresses the condition (1.14) for the perpendicular directions u, v, and $E = G$ expresses this condition for the directions inclined at $\pm 45°$ to u, v.

We may now formulate the problem of Plateau analytically as follows.

Given in n-dimensional euclidean space $\mathfrak{x}$ *any finite number k of sensed contours,* Γ. *To find*

(1) *a Riemann surface R of prescribed topological characteristic r and with k boundaries C,*

[13] Tissot's theorem.

(2) *upon R a harmonic vector $H(Q)$ remaining continuous on the boundaries C,*
such that

(α) *the boundary values of H on C: $\mathfrak{x} = g(P)$, shall form a parametric representation of the given contours,*

(β) *for every point Q of R and every pair of perpendicular directions ξ, η on R at Q,*

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0.$$

This problem solved, the equation

$$\mathfrak{x} = H(Q) \tag{1.15}$$

evidently represents a minimal surface M of characteristic r bounded by the contours Γ.

We may suppose the boundary C of R sensed so that R is on the left; then we require that C shall be converted by the parametric equations (α) into the contours Γ with their preassigned sense of description. Accordingly, M with its normal properly directed, is, by requirement, on the left of Γ.[14]

To fix the ideas, we have given the preceding formulation only for two-sided or orientable surfaces. The adaptation to one-sided or non-orientable surfaces $\bar{M}$, which will be discussed later (§ 2. 5), depends on the use of an orientable covering surface M in two-one point correspondence with $\bar{M}$.

Also, for definiteness, let us suppose that the given contours do not intersect one another—although it is very easy to adapt our method to the case where such intersections are permitted.[15] Under the assumption that the contours Γ have no common points, the boundaries C of R must also be required not to touch one another.

4. The conditions (1. 8), (1.9), or (β), express only the first variational condition in the problem of the surface of least area with the boundaries Γ and of topological type r. Therefore a minimal surface in the sense of these formulas may or may not have least area for its topological form and boundaries. However, the minimal surfaces whose existence is proved in this paper actually do have the least area property. But for the sake of uniformity of subject matter and treatment, we shall not deal with least area here, referring for this to our other publications.[16]

---

[14] A full discussion of the orientation of the contours is given in [9], § 3. There are 2^{k-1} essentially distinct possibilities in this respect.

[15] [5]; [9], § 3. 3.

[16] [2], § 26; [9], §§ 16, 19.

2. PREPARATORY CONSIDERATIONS CONCERNING RIEMANN SURFACES.

1. Concerning Riemann surfaces, we follow the fundamental and highly suggestive tract of F. Klein [17] of about fifty years ago.

Let $\mathfrak{R}$ denote a Riemann surface having an inverse conformal transformation T into itself, where T associates the points of $\mathfrak{R}$ in pairs, so that $T^2 = 1$ (involutory transformation). Such a Riemann surface, following Klein, is called *symmetric*. The locus of points fixed under T consists of a system of closed curves, C, called *curves of transition*. If $\mathfrak{R}$ is considered as the Riemann surface of an algebraic curve $\mathcal{C}$: $P(x, y) = 0$, with *real* coefficients, then T may be taken to be the interchange of conjugate complex points (x, y), $(\bar{x}, \bar{y})$, and the curves of transition are the real branches of the algebraic curve.

The transition curves may separate $\mathfrak{R}$ or not; following Klein, $\mathfrak{R}$ is accordingly said to be of the *first kind* or *second kind*, also *orthosymmetric* or *diasymmetric*. If we regard a pair of T-equivalent or symmetric points as a single geometric element, then $\mathfrak{R}$ becomes a geometric manifold R, sometimes called a *Riemann semi-surface*. However, we shall simply speak of R again as a Riemann surface, while keeping in mind that R has the boundaries C, whereas $\mathfrak{R}$ is closed. An important remark is that R is orientable or non-orientable according as $\mathfrak{R}$ is of the first or second kind. In case $\mathfrak{R}$ is of the first kind, R may simply be identified with one of the two conjugate halves into which $\mathfrak{R}$ is separated by its curves of transition.

Any orientable or non-orientable surface S may be put into correspondence with a symmetry Riemann surface $\mathfrak{R}$ of the first or second kind, respectively, so that a point P of S together with its antipodal $\bar{P}$ correspond to a pair of symmetric points of $\mathfrak{R}$. (The antipodal point to P means the point $\bar{P}$ geometrically coincident with P, but regarded as lying on the opposite side of S, i. e., with reversed sense of the normal.) The distinction between the first and second kind of symmetric Riemann surface corresponds to the fact that on a two-sided surface we cannot pass from a point to its antipodal without crossing the boundary, whereas it is characteristic of a one-sided surface that such a passage is possible.

In the form of the general Riemann semi-surface R, we are provided with a complete system of standard domains of any topological characteristic r, any finite number k of boundaries C, and either one- or two-sided. The completeness is from the standpoint of conformal mapping; i. e., the $3p - 3$ real con-

[17] *Über Riemanns Theorie der Algebraischen Funktionen und ihrer Integrale*, 1882. English translation, Cambridge, 1893.

formal moduli are capable of varying arbirarily in this system. Here ρ denotes the genus of $\mathfrak{R}$, easily seen to be expressed by the formula

$$p = r + k - 1 = 2h + k - 1. \tag{2.1}$$

From our point of view, conformally equivalent Riemann surfaces R or $\mathfrak{R}$ are identical, and one may replace the other if convenient for any purpose. In the case of the complete surface $\mathfrak{R}$, the conformal equivalence must respect symmetry; i. e., convert symmetric points into symmetric points. What is tantamount, the corresponding real algebraic curve $\mathcal{A}$ is subject to real birational transformation without change of anything essential to our discussion.

2. Any Riemann surface $\mathfrak{R}$ may be represented in the standard way as a many-sheeted surface over the complex z-plane. We may even restrict ourselves to the Clebsch-Lüroth normal form,[18] where all branch-points are of first order, i. e., like that pertaining to $\sqrt{z}$. The inverse conformal involution T may always be represented as the *conjugate transformation* $z \to \bar{z}$, *combined with a substitution* S *of period two on the various sheets.* The transition curves C are then branch-cuts along the real axis. If these be slit open, then in the case of $\mathfrak{R}$ of the first kind, the surface separates into two symmetric halves R and R', and R, for instance, is a Riemann surface with the slits C for boundaries. When $\mathfrak{R}$ is of the second kind, we must identify its symmetric points conceptually, whereupon $\mathfrak{R}$ becomes a one-sided manifold R, again having the slits C for boundaries, the opposite edges of which are now to be identified point-wise. In the standard nomenclature, $\mathfrak{R}$ is a two-sided *covering surface* in two-one correspondence with R; i. e., each point of R corresponds to a pair of symmetric points of $\mathfrak{R}$.

3. Illustrations. First, *the one-sided case.* Consider, for example, a double-sheeted Riemann surface $\mathfrak{R}$ over the z-plane. In this, let there be two branch-cuts, one C, or AB, along the real axis, the other $Q\bar{Q}$ joining two conjugate imaginary points. Here the inverse conformal involution T is a reflection in the real axis combined with the interchange $S = (12)$ of the two sheets. The points P, $\bar{P}$ in the figure represent T-equivalent points, if we take P in the upper, $\bar{P}$ in the lower sheet. It is evident that T leaves invariant the points of the branch-cut C, and no other points. C is therefore the unique curve of transition; it is to be regarded as a closed curve, proceeding from A to B in one sheet and back to A in the other.

Obviously, C does not separate $\mathfrak{R}$, for we can pass from P to $\bar{P}$ via the branch-cut $Q\bar{Q}$ without traversing C, as indicated in the figure. If we identify

[18] See E. Picard, *Traité d'Analyse*, vol. 2 (1925), p. 444.

all symmetric point-pairs, the resulting manifold represents a *Möbius surface* bounded by C. Point-pairs such as $P\bar{P}$ correspond to a point on the Möbius surface together with its antipodal, and the one-sided nature of the surface is expressed exactly by the possibility of passing from P to $\bar{P}$ without crossing the boundary C.

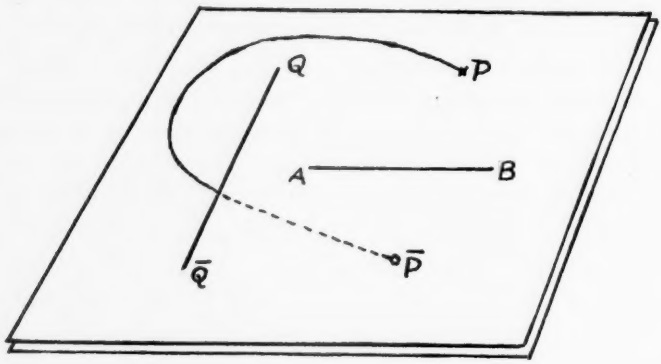

Fig. 1.

By adding more branch-cuts like $Q\bar{Q}$ joining conjugate imaginary points, and branch-cuts like AB along the real axis, we can obtain models for one-sided surfaces of any characteristic r and with any finite number k of boundaries.

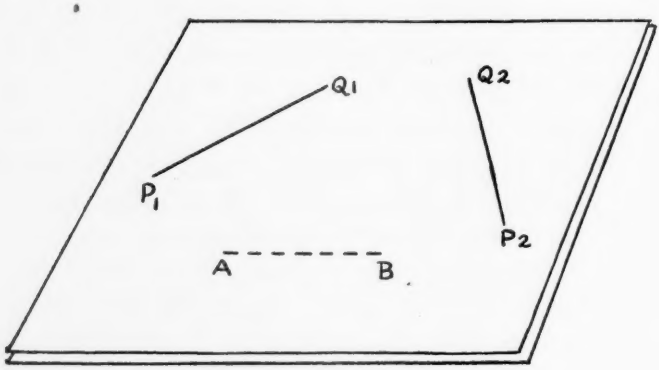

Fig. 2.

Two-sided Riemann surfaces. A typical illustration is the Riemann surface R indicated in fig. 2, which represents a surface topologically equivalent to a torus with a single boundary. Here P_1Q_1, P_2Q_2 are branch-cuts, along which the two sheets are joined. These give the surface its genus $h = 1$,

or characteristic $r = 2$. The dotted line AB, representing the boundary of the surface, is a slit along the real axis in the lower sheet. To obtain the complete symmetric Riemann surface $\mathfrak{R}$, we first reflect the figure in a plane just below the lower sheet, obtaining R', with a slit $A'B'$ in its upper sheet. Then we join opposite edges of AB and $A'B'$ cross-wise, in the usual manner. The resulting Riemann surface $\mathfrak{R}$ has four sheets, and its inverse conformal automorphism T is the conjugate transformation $z \to \bar{z}$ combined with the substitution $S = (14)(23)$, where the sheets are numbered in order from the top. Evidently, the only points fixed under T are those of $AB = A'B'$, which is therefore the unique transition curve of $\mathfrak{R}$.

By adding more branch-cuts like P_1Q_1, P_2Q_2 and more branch-cuts like AB along the real axis in any of the sheets, we can produce Riemann surfaces R with any number k of boundaries and any genus h.

4. *Interpretation in four-dimensional space.* Let $P(x, y) = 0$ be any one of the birationally equivalent class of algebraic equations with real coefficients corresponding to the symmetric Riemann surface $\mathfrak{R}$. Write $x = x_1 + ix_2$, $y = y_1 + iy_2$, and interpret x_1, x_2, y_1, y_2 as coördinates in a four-dimensional euclidean space. Then, when real and imaginary parts are separated, the equation

$$P(x_1 + ix_2, y_1 + iy_2) = 0 \tag{2.2}$$

gives two real equations:

$$P_1(x_1, x_2, y_1, y_2) = 0, \qquad P_2(x_1, x_2, y_1, y_2) = 0 \tag{2.3}$$

and defines therefore a two-dimensional manifold $\mathfrak{S}$ in the four-dimensional space. If we measure angles on $\mathfrak{S}$ according to the metric induced by the surrounding euclidean four-space, then it becomes a riemannian manifold. In fact, it so happens that $\mathfrak{S}$ is exactly a minimal surface in the four-space. For the equation (2.2) is equivalent to the parametric equations

$$x_1 + ix_2 = F(w), \qquad y_1 + iy_2 = G(w), \tag{2.4}$$

$$x_1 = \Re F(w), \qquad x_2 = \Re \frac{1}{i} F(w), \qquad y_1 = \Re G(w) \qquad y_2 = \Re \frac{1}{i} G(w). \tag{2.5}$$

The sum of the squares of the derivatives of the four complex functions which appear here is zero, so that we have a particular case of the Weierstrass equations (1.11, 12). The representation of $\mathfrak{S}$ on the w-domain is conformal, and by the equation (2.4) so are the representations of the $(x = x_1 + ix_2)$-plane and the $(y = y_1 + iy_2)$-plane. It follows that the orthogonal projections of $\mathfrak{S}$ on the complex x- and y-planes are conformal. These projections are Riemann surfaces $\mathfrak{R}_x$, $\mathfrak{R}_y$ of the standard many-sheeted form.

The conformal involution T of $\mathfrak{S}$ consists of the interchange of conjugate complex points, i. e., the transformation of (x_1, x_2, y_1, y_2) into $(x_1, -x_2, y_1, -y_2)$. The real branches of the algebraic curve $\mathcal{A}$, $P(x, y) = 0$, are formed by the section of $\mathfrak{S}$ with the real $x_1 y_1$-plane: $x_2 = 0$, $y_2 = 0$. Evidently the points of these real branches and only these points, are left invariant by T, so that the real branches of $\mathcal{A}$ constitute the transition curves of the symmetric Riemann surface $\mathfrak{S}$. Accordingly—repeating a previous remark in another form—we have two types of real algebraic curve, those whose Riemann manifolds $\mathfrak{S}$ are separated by the real branches, and those which are not thus separated. These correspond respectively to two- and one-sided semi-manifolds S, obtained by identification of symmetric points of $\mathfrak{S}$.

Since conformally equivalent Riemann surfaces are equivalent for all our purposes, we may and shall use $\mathfrak{R} = \mathfrak{R}_x$ and $\mathfrak{S}$ interchangeably, as well as the semi-surfaces R and S.

5. The Plateau problem for one-sided surfaces. We are now in a position to give an exact analytic formulation of the Plateau problem also for one-sided minimal surfaces.[19] *All one-sided surfaces and quantities pertaining to them will be denoted by means of a bar*; thus we shall speak of a surface $\bar{M}$ of topological type $\bar{r}$, meaning a one-sided surface of characteristic r, which now may be either an odd or an even integer.

Let $\mathfrak{R}'$ denote a Riemann surface of the second kind, severed along its transition curves C, and whose semi-surface R, resulting by identification of of symmetric points, is of topological type $\bar{r}$.

Given:

(1) *any k contours Γ in euclidean $\mathfrak{x}$-space of n dimensions,*

(2) *an assigned topological type $\bar{r}$;*

to find:

(1) *a Riemann surface $\mathfrak{R}'$ of the second, or diasymmetric, kind whose semi-surface $\bar{R}$ is of topological type $\bar{r}$.*

(2) *upon $\mathfrak{R}'$ a harmonic vector $H(Q)$ which takes equal values at symmetric points*:

$$(\gamma) \qquad H(\bar{Q}) = H(Q),$$

and which remains continuous on the transition curves C along which $\mathfrak{R}'$ is severed;

---

[19] Cf. [4], where this type of formulation is given for the typical case of a Möbius strip.

such that

(α) *the boundary values on* C *of* H: $\mathfrak{x} = g(P)$, *shall form a parametric representation of* Γ;

(β) *for every point* Q *of* $\mathfrak{R}'$ *and every pair of perpendicular directions* ξ, η,

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0.$$

Then, evidently, the equation $\mathfrak{x} = H(Q)$ represents a one-sided minimal surface $\bar{M}$ bounded by Γ. According to condition (γ), each pair of symmetric points $Q, \bar{Q}$ of $\mathfrak{R}'$ give geometrically coincident points $P, \bar{P}$ of $\bar{M}$ where $\bar{P}$ is antipodal to P. Since we can pass on $\mathfrak{R}'$ from Q to $\bar{Q}$ without traversing C, we can pass on $\bar{M}$ from P to $\bar{P}$ without traversing Γ.

The transition curves C of $\mathfrak{R}'$ may be sensed in an arbitrary way, and these senses of decsription on C can be made to correspond by the equations (α) to preassigned senses of description on the contours Γ.[20]

6. Reduction of Riemann surfaces. The form of the symmetric Riemann surface $\mathfrak{R}$ is completely determined by the number of its sheets, the number and position of its branch-points, and the position of the transition curves C along the real axis of the various sheets. If these vary, then in the limit $\mathfrak{R}$ may take various degenerate or reduced forms, arising from the coalescence of branch-points, i. e., the disappearance of certain branch-cuts. In this way, the semi-surface R may separate into a number of disconnected parts, or its characteristic r may be lowered. In case of separation, the total characteristic of the component parts is evidently $\leqq r$. We shall term either type of degeneration a *reduction* of R.

Another logically possible type of degeneration of R consists in the shrinking to a point of some of the boundaries C. However, we shall see that this can be avoided by conformal or birational transformation of $\mathfrak{R}$. Still another improper form occurs when two transition curves of $\mathfrak{R}$, or real branches of the corresponding algebraic curve $\mathfrak{A}$, touch in the limit, i. e., $\mathfrak{A}$ acquires a *real double-point*. How this is avoided in the last analysis is easily shown in § 8. 8.

Every form of reduction can be obtained by composition of primary reductions. A *primary reduction* of R consists of the following.

(1) In the two-sided case, the coalescence of two branch-points, i. e., the disappearance of a single branch-cut.

(α) If the disappearing branch-cut is the only one which unites two given sheets, then R separates into two disconnected parts: $R = R_1 + R_2$, with

[20] There are 2^{k-1} possibilities in this respect. See footnote 14 and [9], § 3.

no decrease of the total characteristic $r = r_1 + r_2$. The k slits C, forming the boundaries of R, are distributed between R_1 and R_2: $k = k_1 + k_2$. We may have $k_2 = 0$, in which case R_2 is a closed surface, and, as will be obvious, may simply be ignored for our purposes. If R_2 is closed, we shall not call the reduction primary unless $r_2 = 2$;[21] if $r_2 = 0$,[22] it may be considered that no reduction at all has taken place.

(b) If branch-cuts remain which join the same sheets as the disappearing one, then the characteristic of R is reduced by exactly two units, to $r - 2$.

(2) In the one-sided case, $\mathfrak{R}'$, a primary reduction will mean:

(a) The disappearance of a single pair of mutually symmetric branch-cuts, provided that this produces separation of $\mathfrak{R}'$: $\bar{R} = \bar{R}_1 + \bar{R}_2$ or $\bar{R} = \bar{R}_1 + R_2$, with $r = r_1 + r_2$. If a component $\bar{R}_2$ is closed, we require also $r_2 = 1$. As indicated in the notation, at least one of the components must be one-sided, and both may be.

(b) The coalescence of two symmetric branch-points Q, $\bar{Q}$, whose point of coincidence must obviously be real. This reduces the characteristic of R by exactly *one* unit, to $r - 1$. If $r - 1$ is even, the reduced surface may be two-sided or one-sided; if $r - 1$ is odd, it must be one-sided.

We shall denote primary reduction by an accent: r', $\bar{r}'$, and the general type of reduction by r'', $\bar{r}''$, *interpreting the latter as including the former.* A *composite reduction* will mean one which is not primary.

In a composite reduction of a one-sided surface there may be any distribution of one- and two-sided components.

In terms of the corresponding algebraic equation $P(x, y) = 0$, the separation of R is equivalent to reducibility in the usual algebraic sense:

$$P(x, y) = P_1(x, y)\, P_2(x, y) \cdots P_m(x, y). \tag{2.6}$$

The lowering of the genus of R corresponds to the acquisition of new conjugate imaginary double points, produced by the coalescence of branch-points.

7. An example. According to the formula (2.1), $p = 2h + k - 1$, a Riemann surface R with two boundaries and of genus $h = 0$ corresponds to a real algebraic curve of genus $p = 1$ with two real branches. As has been said, such a curve can be realized in many different ways, for we can subject any given one to the infinite group of real birational transformations. An example of a complete family of curves obeying the stated conditions is

$$(x^2 + y^2 - 1)(x - 2) = m, \tag{2.8}$$

[21] The detachment of a closed component of genus one.

[22] The detachment of a closed component of genus zero.

where

$$0 < m < m_0 = \left(\frac{-2+\sqrt{28}}{6}\right)\left(\frac{10+\sqrt{28}}{6}\right)\left(\frac{8-\sqrt{28}}{6}\right).$$

The completeness is from the standpoint of birational or conformal transformation with preservation of symmetric points.

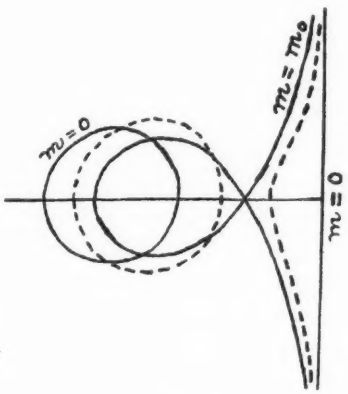

Fig. 3.

When $m = 0$, the curve is reducible, separating into a circle and straight line. When $m = m_0$, the two real branches come together, producing a real double point. The figure shows the extreme curves $m = 0$, $m = m_0$ drawn full, and an intermediate one drawn dotted. If desired, the presence of the infinite branch could be obviated by means of an inversion.

Another complete family of real algebraic curves of genus one with two real branches is represented by

$$y^2 = x(x-1)(x-2)(x-m), \qquad (0 < m < 1). \tag{2.8}$$

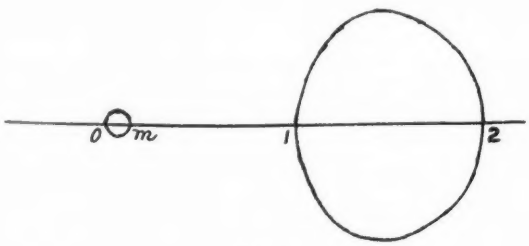

Fig. 4.

Here, if $m = 1$, the curve acquires a real double point. If $m = 0$, one of the real branches shrinks to a point. But by a real birational transformation we can

then pass to the algebraic curves of fig. 3, where the corresponding singularity is reducibility of the algebraic equation or separation of the Riemann surface.

This mode of disappearance of a real branch of the algebraic curve $\mathcal{A}$ is quite a troublesome kind of singularity in our analysis. For we use each of these real branches on which to represent a contour Γ parametrically. The preceding example illustrates, in a sufficiently typical way, how we can always replace this troublesome type of singularity by the much more easily handled one of separation of the Riemann surface.

3. FORMULAS.

1. Let $H(Q)$ denote the harmonic vector function on the Riemannn surface R which has the values $g(P)$ on the boundaries C. Then, in terms of Green's function $G(P, Q)$ of R, the function $H(Q)$ is given by the classic formula

$$H(Q) = \frac{1}{2\pi} \int_C g(P) \frac{\partial G(P, Q)}{\partial n} \, ds. \tag{3.1}$$

Here $\partial/\partial n$ refers to the interior normal to C at the point P.

Green's function $G(P, Q)$ of R is defined uniquely by the following properties. As a function of the point P, $G(P, Q)$ is:

(I) *uniform* on R,

(II) *harmonic* on R,

(III) regular except for a logarithmic singularity at the point Q, in whose neighborhood,

$$G(P, Q) = \log \frac{1}{\overline{PQ}} + G_1(P, Q),$$

where $\overline{PQ}$ = distance PQ, and $G_1(P, Q)$ is regular at the point Q;

(IV) equal to zero if P is on the boundary C.

The existence of Green's function for any Riemann surface may be assumed as a classical result. An explicit formula in terms of θ-functions and abelian integrals has recently been given by the author in [9, 10], and is reproduced here as (7. 39), with reference to (7. 7).[23]

We may recall the well-known symmetry property of Green's function:

$$G(P, Q) = G(Q, P). \tag{3.1'}$$

[23] The case $h = 0$ of this formula was given in [8], as formula (2).

2. Let ξ, η denote any two perpendicular directions at the arbitrary point Q of R. Then we may differentiate (2.1) in the directions ξ, η under the integral sign, and get

$$\frac{\partial H(Q)}{\partial \xi} = \frac{1}{2\pi}\int_C g(P_1)\,\frac{\partial^2 G(P_1, Q)}{\partial n_1 \partial \xi}\,ds_1, \tag{3.2}$$

$$\frac{\partial H(Q)}{\partial \eta} = \frac{1}{2\pi}\int_C g(P_2)\,\frac{\partial^2 G(P_2, Q)}{\partial n_2 \partial \eta}\,ds_2. \tag{3.3}$$

By forming the scalar product of these two vectors, we have

$$\frac{\partial H(Q)}{\partial \xi}\cdot\frac{\partial H(Q)}{\partial \eta} = \frac{1}{4\pi^2}\int_C\int_C g(P_1)\cdot g(P_2)\,\frac{\partial^2 G(P_1, Q)}{\partial n_1 \partial \xi}\,\frac{\partial^2 G(P_2, Q)}{\partial n_2 \partial \eta}\,ds_1 ds_2. \tag{3.4}$$

Since the scalar product is commutative, we may interchange ξ, η and obtain

$$\frac{\partial H(Q)}{\partial \xi}\cdot\frac{\partial H(Q)}{\partial \eta} = \frac{1}{4\pi^2}\int_C\int_C g(P_1)\cdot g(P_2)\,\frac{\partial^2 G(P_1, Q)}{\partial n_1 \partial \eta}\,\frac{\partial^2 G(P_2, Q)}{\partial n_2 \partial \xi}\,ds_1 ds_2. \tag{3.5}$$

Taking half the sum of the last two formulas, we find the more symmetric expression

$$\frac{\partial H(Q)}{\partial \xi}\cdot\frac{\partial H(Q)}{\partial \eta} = \frac{1}{8\pi^2}\int_C\int_C g(P_1)\cdot g(P_2)\,\frac{\partial^2}{\partial n_1 \partial n_2}\left\{\frac{\partial G(P_1, Q)}{\partial \xi}\,\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta}\,\frac{\partial G(P_2, Q)}{\partial \xi}\right\} ds_1 ds_2. \tag{3.6}$$

We may replace the factor $g(P_1)\cdot g(P_2)$ by a geometrically more significant one, as follows:

$$\frac{\partial H(Q)}{\partial \xi}\cdot\frac{\partial H(Q)}{\partial \eta} = -\frac{1}{16\pi^2}\int_C\int_C [g(P_1) - g(P_2)]^2\,\frac{\partial^2}{\partial n_1 \partial n_2}\left\{\frac{\partial G(P_1, Q)}{\partial \xi}\,\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta}\,\frac{\partial G(P_2, Q)}{\partial \xi}\right\} ds_1 ds_2. \tag{3.7}$$

Each of the added terms has a factor of the type

$$\int_C \frac{\partial^2 G(P_2, Q)}{\partial n_2 \partial \eta}\,ds_2 = \frac{\partial}{\partial \eta}\int_C \frac{\partial G(P_2, Q)}{\partial n_2}\,ds_2 = 0, \tag{3.8}$$

for according to (3.1) with $H(Q) \equiv 1 \equiv g(P)$, the second integral has the constant value 2π. Therefore the addition of these terms, while changing the form of the expression advantageously, has no effect on its value.

3. Formula for $A(g, R)$. We have defined

$$A(g, R) = D(H) = \frac{1}{2}\int\int_R (H_u^2 + H_v^2)\, dudu. \tag{3.9}$$

Since the behavior of the partial derivatives of H as we approach the boundary C of R is uncertain, this integral is improper, and must be considered as the limit of

$$A'(g, R) = \frac{1}{2}\int\int_{R'} (H_u^2 + H_v^2)\, dudv, \tag{3.10}$$

where R', inclusive of its boundary C', is an interior sub-region of R, and $R' \to R$, $C' \to C$.

By Green's formula, (3.10) transforms into an integral over C':

$$A'(g, R) = -\frac{1}{2}\int_{C'} H(Q) \cdot \frac{\partial H(Q)}{\partial n'}\, ds', \tag{3.11}$$

where $\partial/\partial n'$ is in the direction of the interior normal to C'.

Substituting from (3.2), we obtain

$$A'(g, R) = -\frac{1}{4\pi}\int_C \int_{C'} g(P) \cdot H(Q)\, \frac{\partial^2 G(P, Q)}{\partial n \partial n'}\, dsds'. \tag{3.12}$$

By a remark similar to that associated with (3.8), this can be transformed without change of value into

$$A'(g, R) = \frac{1}{8\pi}\int_C \int_{C'} [g(P) - H(Q)]^2\, \frac{\partial^2 G(P, Q)}{\partial n \partial n'}\, dsds'. \tag{3.13}$$

Let now $C' \to C$; then the arbitrary point Q of C' approaches to the arbitrary point P_2 of C, and $H(Q)$ tends to $g(P_2)$. Therefore the preceding formula becomes in the limit:

$$A(g, R) = \frac{1}{8\pi}\int_C \int_C [g(P_1) - g(P_2)]^2\, \frac{\partial^2 G(P_1, P_2)}{\partial n_1 \partial n_2}\, ds_1 ds_2. \tag{3.14}$$

We shall not stop for the entirely routine justification of this passage to the limit under the integral sign. A complete discussion has been given for the sufficiently typical one-contour case.[24]

---

[24] [2], § 21.

We observe the following interpretation of the "kernel" $\partial^2 G(P_1, P_2)/\partial n_1 \partial n_2$. Let Q_1, Q_2 denote, respectively, the points in the interior of R obtained by proceeding the elementary distances dn_1, dn_2 in the direction of the normals to C at P_1, P_2. Then the second normal derivative can be expressed as limit of a second difference quotient, as follows:

$$(3.15)\quad \frac{\partial^2 G(P_1, P_2)}{\partial n_1 \partial n_2} = \lim_{dn_1, dn_2 \to 0} \frac{G(Q_1, Q_2) - G(Q_1, P_2) - G(P_1, Q_2) + G(P_1, P_2)}{dn_1 dn_2} = \lim_{dn_1, dn_2 \to 0} \frac{G(Q_1, Q_2)}{dn_1 dn_2},$$

by property (IV) of Green's function.

But for any two interior points Q_1, Q_{21} it is known that $G(Q_1, Q_2)$ is always *positive*. In fact, the range of values of $G(P, Q)$ as P ranges over the interior of R is from 0 at the boundary to $+\infty$ at the pole Q. *Hence the kernel*

$$(3.16)\quad G''(P_1, P_2) = \frac{\partial^2 G(P_1, P_2)}{\partial n_1 \partial n_2}$$

of $A(g, R)$ is always positive.

Further, we see that $G''(P_1, P_2)$ has a singularity when $P_1 = P_2$ with principal term

$$(3.17)\quad \frac{2}{\overline{P_1 P_2}^2}.$$

This results from the logarithmic singularity of $G(P_1, P_2)$ according to property (III). It follows that the formula (3.14) for $A(g, R)$ is an *improper* integral, and must be evaluated by integrating first with $\overline{P_1 P_2} \geqq \epsilon$, and then allowing $\epsilon \to 0$. On account of the positive nature of the integrand, $A(g, R)$ always exists definitely either as a finite quantity or as $+\infty$.

Logically, we can *begin* with the formula (3.14) as the *definition* of $A(g, R)$, without regard to the equivalence to the vector Dirichlet integral $D(H)$. Therefore, *the proof of this paper does not use the minimum property of the Dirichlet integral,*[25] i. e.,

$$(3.18)\quad D(H) \leqq D(\mathfrak{x}),$$

where $\mathfrak{x}(u, v)$ is any piece-wise continuously differentiable vector with the same boundary values as the harmonic vector $H(u, v)$.

[25] The same remark applies to the author's original paper [2], and to the others contained in the list at the end of the present one.

We may observe, in passing, that the formula (3. 14) for $D(H)$ applies to any number m of dimensions of the domain R, the integral being taken over SS, the geometric product with itself of the $(m-1)$-dimensional boundary S of R. The only change is in the numerical factor $1/8\pi$ to correspond to the dimensionality m.

3. The one-sided case. Let $\mathfrak{R}$ denote a symmetric Riemann surface of the second kind, on whose k transition curves C a continuous function $g(P)$ is given. Suppose $\mathfrak{R}$ to be slit along its transition curves; then it becomes a Riemann surface $\mathfrak{R}'$ with $2k$ boundaries, geometrically coincident in pairs point by point, but conceptually distinct. The boundary function $g(P)$ now defines a Dirichlet problem on $\mathfrak{R}'$ whose solution is a harmonic function $H(Q)$ coinciding with $g(P)$ on C.

Let us construct the function $\tilde{H}(Q) = H(\bar{Q})$. Then, since the transformation T of Q into $\bar{Q}$ is inversely conformal, it follows that $\tilde{H}(Q)$ is harmonic. Also, since for any point P on C we have $\bar{P} = TP = P$, the boundary values of $\tilde{H}(Q)$ evidently coincide with those of $H(Q)$. Hence $\tilde{H}(Q) \equiv H(Q)$; i. e.,

$$H(Q) = H(\bar{Q}). \tag{3. 19}$$

Thus the condition (γ) of the formulation of § 2. 5 is satisfied automatically whenever $g(P)$ takes equal values at geometrically coincident points of the slits or boundaries of $\mathfrak{R}'$.

We suppose this condition always verified, and define $A(g, R)$ in the one-sided case by the formula

$$A(g, R) = \tfrac{1}{2}D(H) = \tfrac{1}{4}\int\int_{\mathfrak{R}'}(H_u^2 + H_v^2)\, dudv. \tag{3. 20}$$

The extra factor factor 1/2, as compared with (3. 14), is intended to neutralize the circumstance that each element of the surface $\mathfrak{x} = H(u, v)$ is obtained twice, with opposite senses, corresponding to T-equivalent or symmetric elements of $\mathfrak{R}'$.

We may repeat that the Green's function that appears in this formula is relative to $\mathfrak{R}'$ as a manifold with $2k$ boundaries, produced by the severance of its transition curves. The reader will find it easy to illustrate by means of figure 1 and analogous figures with more branch-cuts like AB and $Q\bar{Q}$.

4. OUTLINE OF THE SOLUTION.

In the functional $A(g, R)$ let us substitute for R all Riemann surfaces of topological type r or $\bar{r}$ and with k boundaries, and for g all possible repre-

sentations of the given contours Γ on the boundary C of R: $\mathfrak{x} = g(P)$. To obtain a *closed* set $[(g, R)]$, we must adjoin also various types of improper representations, where R is reduced, or where g is defined by an improper monotonic correspondence between Γ and C. Then, on account of the *lower semi-continuity* of the functional $A(g, R)$ and the *compactness* of its argument range $[(g, R)]$, it will be merely a routine matter to prove the *attainment* of the minimum of $A(g, R)$:

$$A(g^*, R^*) = \text{min.}; \tag{4.1}$$

the procedure follows the standard Weierstrass-Fréchet pattern, which we have presented many times before in previous papers.[26] Certain natural sufficient conditions (8.22, 23) then preclude the possibility that R^* should reduce or that g^* should be an improper parametric representation.

Thus we come to the variational condition:

$$\delta A(g, R) = 0 \tag{4.2}$$

for $(g, R) = (g^*, R^*)$. It is the treatment of this variation of $A(g, R)$ which constitutes the central theme of the present paper.

Suppose the Riemann surface R to vary in dependence on a parameter ϵ, taking the series of forms R_ϵ, where for $\epsilon = 0$ R_ϵ is R^*. We may imagine this variation to take place by displacement of the individual points $P(\epsilon)$ of R_ϵ. Suppose that the correspondence of $P(\epsilon)$ on R_ϵ to P on R^* gives a one-one and conformal relation between R_ϵ and R^*, at least in the neighborhood of the boundaries C_ϵ, C^*. Then certainly C_ϵ corresponds to C^* in a topological way, so that the parametric representation g^* of Γ on C^* is carried over into a parametric representation g_ϵ on C_ϵ.

Thus we have a *variation* (g_ϵ, R_ϵ) *of the representation* (g^*, R^*) of the contours Γ.[27] Then we may write

$$A(g_\epsilon, R_\epsilon) = J(\epsilon), \tag{4.3}$$

and by $\delta A(g, R)$ we shall understand

$$\delta A(g, R) = J'(0). \tag{4.4}$$

Since the position of the points $P_1(\epsilon)$, $P_2(\epsilon)$ on R_ϵ, as well as the form of R_ϵ, depends on ϵ, we may write

---

[26] See especially [6].

[27] The following discussion of the first variation makes no use of the *minimum* character of (g^*, R^*), but is valid for any particular representation (g, R). However, the notation (g^*, R^*) suggests the chief application, which is to prove (4.2) as a consequence of (4.1).

$$G_\epsilon(P_1(\epsilon), P_2(\epsilon)) = \mathfrak{G}(P_1, P_2; \epsilon), \tag{4.5}$$

where G_ϵ is with respect to R_ϵ; and we define

$$\delta G(P_1, P_2) = \frac{\partial}{\partial \epsilon} \mathfrak{G}(P_1, P_2; \epsilon)|_{\epsilon=0} = \mathfrak{G}'(P_1, P_2; 0), \tag{4.6}$$

P_1, P_2 denoting any fixed points of R^*.

Let Q denote an *arbitrary* interior point of R^*, and ξ, η *any* two perpendicular directions on R^* through Q. Then the essence of our proof consists in the following fact.

It is possible to construct a particular variation R_ϵ such that

$$\delta G(P_1, P_2) = \frac{\partial G(P_1, Q)}{\partial \xi} \frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta} \frac{\partial G(P_2, Q)}{\partial \xi}. \tag{4.7}$$

We shall, indeed, give *an explicit construction in both geometric and analytic form* for a variation that accords with this formula.

Suppose, now, that (4.7) has been established. Then let dn_1, dn_2 denote the normal elements to C^* at P_1, P_2, equal respectively to the tangential elements ds_1, ds_2. We thus have the equality of operators:

$$ds_1 ds_2 \frac{\partial^2}{\partial n_1 \partial n_2} = dn_1 dn_2 \frac{\partial^2}{\partial n_1 \partial n_2}. \tag{4.8}$$

According to formula (3.15), this is equivalent to taking the second difference of $G(P_1, P_2)$ when P_1, P_2 undergo the normal displacements dn_1, dn_2. The formation of this second difference is a linear operation, *which is commutative with* δ, according to standard facts of the calculus of variations. Furthermore, the variation of R^* is supposed to be *conformal* in the region adjacent to C; therefore the conditions

$$ds_1 = dn_1, \qquad ds_2 = dn_2 \tag{4.9}$$

remain invariant. It follows that *the operator*

$$ds_1 ds_2 \frac{\partial^2}{\partial n_1 \partial n_2} \tag{4.10}$$

is commutative with δ.

Hence, applying this to (4.7), we have

$$\delta \left\{ \frac{\partial^2 G(P_1, P_2)}{dn_1 dn_2} ds_1 ds_2 \right\} = ds_1 ds_2 \frac{\partial^2}{\partial n_1 \partial n_2} \left\{ \frac{\partial G(P_1, Q)}{\partial \xi} \frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta} \frac{\partial G(P_2, Q)}{\partial \xi} \right\}. \tag{4.11}$$

As standard function-theoretic considerations show, the application of δ to the integral $A(g, R)$ (3.14) can be accomplished by performing this operation under the sign of integration on the "elementary kernel"

$$\frac{\partial^2 G(P_1, P_2)}{\partial n_1 \partial n_2} ds_1 ds. \tag{4.12}$$

Thus, using (4.11) and comparing with (3.7), we obtain the result:

$$\delta A(g, R) = -2\pi \frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta}. \tag{4.13}$$

Consequently, the variational condition (4.2) implies the condition for a minimal surface:

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0. \tag{4.14}$$

Thus, the proof will be completed.

The next section is occupied with the construction of the special variation (4.7).

5. VARIATION OF GREEN'S FUNCTION.

1. Let $\mathcal{A}$ denote any one of the class of birationally equivalent real algebraic curves belonging to the symmetric Riemann surface $\mathcal{R}$. Let Q denote any point of $\mathcal{R}$ not on a transition curve, and $\bar{Q}$ its symmetric point; these correspond on $\mathcal{A}$ to two *conjugate imaginary* points. The tangent lines t and $\bar{t}$ to $\mathcal{A}$ at Q and $\bar{Q}$, respectively, intersect in a real point O. Let the plane containing $\mathcal{A}$ be subjected to a real projective transformation that sends the point O to the point Y, located at infinity in the direction of the y-axis. Then $\mathcal{A}$ becomes a birationally equivalent real algebraic curve $\mathcal{A}_1$, on which $Q_1, \bar{Q}_1$, the images of $Q, \bar{Q}$, are evidently symmetrically located *branch-points.*[28] We may suppose these to be of first order, i. e., of the type pertaining to $\sqrt{z}$, since if the tangent lines $t, \bar{t}$ had higher than the first order of contact with $\mathcal{A}$ at $Q, \bar{Q}$, this could be obviated by a preliminary birational transformation (e. g., inversion in a circle).

Since real birational transformation changes nothing essential, no generality is lost, then, if we consider from the beginning a symmetric Riemann surface $\mathcal{R}$ in the form of a many-sheeted surface over the complex x-plane with branch-points at $Q, \bar{Q}$ and elsewhere.

Let Q and $\bar{Q}$ be subjected to *symmetric* displacements, *while all other*

[28] For, geometrically, a branch-point is one where the tangent line is parallel to the y-axis.

branch-points of $\mathfrak{R}$ stay fixed. Then $\mathfrak{R}$ undergoes a change of form to $\mathfrak{R}_\epsilon$, remaining symmetric. If we associate with each point P of $\mathfrak{R}$ the point $P(\epsilon)$ of $\mathfrak{R}_\epsilon$ geometrically coincident with P in the same sheet—thus belonging to the same value of x, then we have the type of variation of $\mathfrak{R}$ described in the preceding section. Indeed, this variation is *one-one, continuous, and conformal,* except in the immediate vicinity of the moving branch-points Q, $\bar{Q}$, and also respects symmetry—therefore converting the transition curves C of $\mathfrak{R}$ into those of $\mathfrak{R}_\epsilon$.

Since $\mathfrak{R}_\epsilon$ may be rotated through any angle without change of anything essential, we may suppose, without harm to generality, that the direction in which Q is displaced is that of pure imaginaries; i. e., that the branch-point x_0 at Q moves according to the formula

$$x_0(\epsilon) = x_0(Q) + i\epsilon. \tag{5.1}$$

Our problem is to determine the effect of this variation on the Green's function $G(P_1, P_2)$ of the semi-surface R, particularly to find (repeating formula 4. 6)

$$\delta G(P_1, P_2) = \frac{\partial}{\partial \epsilon} G_\epsilon(P_1(\epsilon), P_2(\epsilon))\,|_{\epsilon=0} \tag{5.2}$$

$$= \frac{\partial}{\partial \epsilon} \mathfrak{G}(P_1, P_2; \epsilon)\,|_{\epsilon=0} = \mathfrak{G}'(P_1, P_2; 0),$$

where G_ϵ refers to R_ϵ.

To this end, we shall consider P_2 as an *arbitrary* point of R *remaining fixed throughout the proof,* and suppose P_1 to be the *primary variable* in $G(P_1, P_2)$, $\mathfrak{G}(P_1, P_2; \epsilon)$, $\mathfrak{G}'(P_1, P_2; 0)$.

First, $\mathfrak{G}'(P_1, P_2; 0)$ is a *harmonic* function, since Laplace's equation

$$\frac{\partial^2 \mathfrak{G}(P_1, P_2; \epsilon)}{\partial x'^2} + \frac{\partial^2 \mathfrak{G}(P_1, P_2; \epsilon)}{\partial x''^2} = 0, \tag{5.3}$$

being satisfied identically in ϵ, is evidently preserved under the operator δ. Here $x = x' + ix''$ is the affix of the point P_1.

Second, since for P_1 on C we have, by property (IV) of Green's function,

$$\mathfrak{G}(P_1, P_2; \epsilon) = 0 \tag{5.4}$$

identically in ϵ, it follows similarly that for P_1 on C:

$$\mathfrak{G}'(P_1, P_2; 0) = 0. \tag{5.5}$$

Being a harmonic function, $\mathfrak{G}(P_1, P_2; \epsilon)$ can be expressed as the real part of an analytic function of the complex variable x:

$$\mathfrak{G}(P_1, P_2; \epsilon) = \Re\phi(x; \epsilon). \tag{5.6}$$

In the neighborhood of any point x_0 of the Riemann surface R *not coinciding with the pole* P_2 *of* $G(P_1, P_2)$, the function ϕ can be expressed in the form

$$\phi(x;\epsilon) = c_0(\epsilon) + c_1(\epsilon)t + c_2(\epsilon)t^2 + \cdots, \tag{5.7}$$

where t is the *local uniformizing variable* or *local parameter*: $t = x - x_0$ if x_0 is an ordinary point, $t = 1/x$ if x_0 is at infinity, and $t = (x - x_0)^{1/m}$ if x_0 is a branch-point of order $m - 1$, where, as has been stated, it is sufficiently general to suppose $m = 2$.

On the other hand, *if* x_0 *is the pole* P_2 *of* $G(P_1, P_2)$, we have, according to the defining property (III) of Green's function,

$$\phi(x;\epsilon) = -\log t + c_0(\epsilon) + c_1(\epsilon)t + c_2(\epsilon)t^2 + \cdots. \tag{5.8}$$

Provided that x_0 is not coincident with the branch-point Q, *the local parameter* $t = x - x_0$ *or* $(x - x_0)^{1/2}$ *is independent of* ϵ, and we have therefore for *either* (5.7) *or* (5.8):

$$\frac{\partial}{\partial\epsilon}\phi(x;\epsilon)\,|_{\epsilon=0} = c_0'(0) + c_1'(0)t + c_2'(0)t^2 + \cdots. \tag{5.9}$$

Hence

$$\mathfrak{G}'(P_1, P_2; 0) = \mathfrak{R}\,\frac{\partial}{\partial\epsilon}\phi(x;\epsilon)|_{\epsilon=0} \tag{5.10}$$

is *regular* about every point of R except Q.

About Q, on the contrary, we have the expansion (5.7), where

$$t = (x - x_0)^{1/2} \tag{5.11}$$

evidently *depends on* ϵ, since x_0 which locates Q varies with ϵ according to (5.1). We have, in fact,

$$\left.\frac{\partial t}{\partial\epsilon}\right|_{\epsilon=0} = -\tfrac{1}{2}(x - x_0)^{-1/2}\left(\frac{dx_0}{d\epsilon}\right)_{\epsilon=0} = -\frac{i}{2t}. \tag{5.12}$$

Accordingly, we now get from (5.7):

$$\frac{\partial}{\partial\epsilon}\phi(x;\epsilon)\,|_{\epsilon=0} = -\frac{i}{2}c_1(0)\,\frac{1}{t} + c_0'(0) + \cdots, \tag{5.13}$$

where the dots stand for a convergent power-series.

Put $t = \xi + i\eta$; then

$$\frac{1}{t} = \frac{\xi - i\eta}{\xi^2 + \eta^2}. \tag{5.14}$$

Let us write for the complex quantity $c_1(0)$:

$$c_1(0) = \gamma_1 + i\gamma_2. \tag{5.15}$$

Then, by (5.10), (5.13), the singular part of $\mathcal{G}'(P_1, P_2; 0)$ at $P_1 = Q$ is

$$\frac{1}{2}\ \frac{\gamma_2\xi - \gamma_1\eta}{\xi^2 + \eta^2}. \tag{5.16}$$

We have, referring to (5.6), (5.7),[29]

$$c_1(0) = \frac{d\phi(t;0)}{dt}\bigg|_{t=0} = \frac{\partial G(Q, P_2)}{\partial \xi} - i\,\frac{\partial G(Q, P_2)}{\partial \eta}; \tag{5.17}$$

so that

$$\gamma_1 = \frac{\partial G(P_2, Q)}{\partial \xi}, \qquad \gamma_2 = -\frac{\partial G(P_2, Q)}{\partial \eta}, \tag{5.18}$$

with use of the symmetric property (3.1′) of Green's function. By substitution in (5.16), the singular part of $\mathcal{G}'(P_1, P_2; 0)$ for $P_1 = Q$ is seen to be

$$-\tfrac{1}{2}\left\{\frac{\xi^2 + \eta^2}{\xi}\,\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\eta}{\xi^2 + \eta^2}\,\frac{\partial G(P_2, Q)}{\partial \xi}\right\}. \tag{5.19}$$

Let us now compare with

$$\Delta(P_1, P_2) = \frac{\partial G(P_1, Q)}{\partial \xi}\,\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta}\,\frac{\partial G(P_2, Q)}{\partial \xi}. \tag{5.20}$$

As function of P_1, $G(P_1, Q)$ has at Q the singular part

$$-\log|t| = -\tfrac{1}{2}\log(\xi^2 + \eta^2). \tag{5.21}$$

Hence, denoting by S. P. the singular part at Q, we have

$$\text{S. P.}\,\frac{\partial G(P_1, Q)}{\partial \xi} = \frac{\xi}{\xi^2 + \eta^2}, \qquad \text{S. P.}\,\frac{\partial G(P_1, Q)}{\partial \eta} = \frac{\eta}{\xi^2 + \eta^2}. \tag{5.22}$$

In deriving this, we observe that $-\xi$, $-\eta$ are the coördinates of Q relative to P_1 in the domain of the local parameter t; hence the change of sign as compared with the straightforward derivative.

By reference to (5.19), (5.20), we see that $-\frac{1}{2}\Delta(P_1, P_2)$ *and* $\mathcal{G}'(P_1, P_2; 0)$ *have the same singular part at* Q.

Also, $\Delta(P_1, P_2)$ is evidently *regular everywhere else on* R, and *vanishes on the boundary* C. For if P_1 is on C, then

$$G(P_1, Q) = 0 \tag{5.23}$$

identically in Q; therefore the first factor in each term of $\Delta(P_1, P_2)$ is zero.

Consequently, the difference

[29] If $G(\xi, \eta) = \Re\phi(\xi + i\eta)$, then $\phi'(\xi + i\eta) = \partial G/\partial\xi - i\,\partial G/\partial\eta$.

$$-\tfrac{1}{2}\Delta(P_1, P_2) - \mathfrak{G}'(P_1, P_2; 0)$$

is a harmonic function of P, *regular everywhere on* R *and vanishing on its boundary* C; therefore this function is *identically zero*; i. e.,

$$\mathfrak{G}'(P_1, P_2; 0) = -\tfrac{1}{2}\Delta(P_1, P_2). \tag{5.24}$$

Changing the parameter ϵ to $-\frac{1}{2}\epsilon$, and referring to (5.20) and the notation (4.6), we write this relation in full as

$$\delta G(P_1, P_2) = \frac{\partial G(P_1, Q)}{\partial \xi}\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta}\frac{\partial G(P_2, Q)}{\partial \xi}. \tag{5.25}$$

2. It remains to interpret the directions ξ, η in a geometric way. Geometrically, the branch-point Q is simply the point of contact of the tangent to $\mathcal{A}$ from the infinite point Y of the y-axis. Hence the variation (5.1), $x_0(\epsilon) = x_0(Q) + i\epsilon$, corresponds to a change of form of $\mathcal{A}$ to $\mathcal{A}_\epsilon$ so that the tangent line t_ϵ from Y to $\mathcal{A}_\epsilon$ touches this curve at the new branch-point [30] $Q_\epsilon \sim x_0(\epsilon)$. t_ϵ intersects $\mathcal{A}$ in *two* points $Q'_\epsilon, Q''_\epsilon$ near to Q,[31] which—since they belong to the same value $x_0(\epsilon)$ of x—lie one directly over the other in two different sheets of the x-Riemann surface $\mathfrak{R}$ of $\mathcal{A}$. As ϵ passes through zero from positive to negative, the points $Q'_\epsilon, Q''_\epsilon$ pass through Q, entering from the directions $90°$, $90° + 360°$, and leaving in the directions $-90°$, $-90° - 360°$, respectively.

When we pass to the plane of the local parameter $t = \xi + i\eta = (x - x_0)^{\frac{1}{2}}$, and write $x - x_0 = re^{i\theta}$, $t = \rho e^{i\phi}$, we have $\phi = \frac{1}{2}\theta$. Thus the superposed directions $90°$, $90° + 360°$ become in the t-plane the *opposite* directions $\alpha' = 45°$, $\alpha'' = 45° + 180°$; and $-90°$, $-90° - 360°$ become the opposite directions $\beta' = -45°$, $\beta'' = -45° - 180°$, *perpendicular* to the preceding. The axial directions ξ, η are evidently the *angle-bisectors* of $\alpha'\beta'$, $\alpha'\beta''$.

To be emphasized is the fact that all these relations are invariant under conformal transformation of $\mathfrak{R}$, *or birational transformation of* $\mathcal{A}$, *and are therefore independent of the particular normal forms which have been chosen for purposes of simplification.*

Since the branch-points other than Q, $\bar{Q}$ *do not move*, the tangent lines from Y to $\mathcal{A}$ other than YQ, $Y\bar{Q}$ *remain invariant* when $\mathcal{A}$ changes to $\mathcal{A}_\epsilon$.

Furthermore, we have set the Riemann surfaces R_ϵ and R into corre-

[30] The notation $\sim$ represents that $x_0(\epsilon)$ is the orthogonal projection on the x-Riemann surface of $\mathcal{A}$ of the point Q_ϵ of $\mathfrak{F}$, which is $\mathcal{A}$ regarded as a complete manifold of complex points.

[31] See footnote 35.

spondence by equal values of x. This is evidently equivalent to projecting $\mathcal{A}_\epsilon$ on $\mathcal{A}$ by straight lines from Y.

We may now go back to the original form of $\mathcal{A}$, where Y corresponds to O, the real intersection point of the conjugate imaginary tangents at $Q, \bar{Q}$. Then $\mathcal{A}$ and $\mathcal{A}_\epsilon$ are now set into correspondence by rectilinear projection from O, and all tangents from O to $\mathcal{A}_\epsilon$ except $t_\epsilon = OQ$, $\bar{t}_\epsilon = O\bar{Q}$ remain invariant, in particular, *all real tangents*. We thus have a figure like the following.

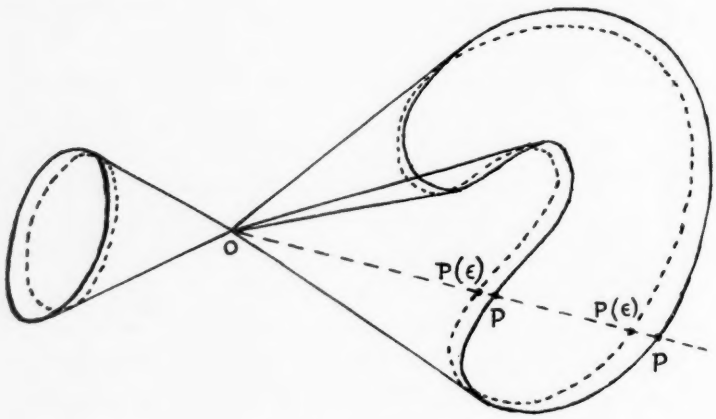

Fig. 5.

$\mathcal{A}$ is drawn full and $\mathcal{A}_\epsilon$ dotted. The correspondence established between the Riemann surfaces of $\mathcal{A}$ and $\mathcal{A}_\epsilon$ by the projection from O is one-one, continuous, and conformal, except in the immediate vicinity of the points $Q, \bar{Q}$; also *the real branches of* $\mathcal{A}$ *and* $\mathcal{A}_\epsilon$ *correspond.* This is because these properties obviously hold for the normal form previously described, where the correspondence between the Riemann surfaces is by equal values of x—therefore actually a congruent transformation, except in the neighborhood of the moving branch-points $Q, \bar{Q}$. It is evident from the figure how the circumstances that the real branches of $\mathcal{A}$ and $\mathcal{A}_\epsilon$ are in one-one continuous correspondence depends on the invariance of the real tangents from O. For if a real arc of $\mathcal{A}_\epsilon$, for instance, protrudes beyond a real tangent to $\mathcal{A}$ then obviously the points of this arc can have no real corresponding points on $\mathcal{A}$.

In the interpretation of these conditions, it is well to have clearly in mind the different equivalent forms $\mathcal{S}$ and $\mathcal{R} = \mathcal{R}_x$ of the Riemann surface of $\mathcal{A}$.[32]

We can easily express the variation of the algebraic curve in explicit form

[32] Cf. § 2. 4.

as follows. Let us employ homogeneous trilinear coördinates (x, y, z) in the plane, as well as the dual line coördinates (u, v, w):

$$ux + vy + wz = 0. \tag{5.26}$$

Taking O as one vertex, $w = 0$, of the triangle of reference, we see that the *line equation* of $\mathcal{A}$ must have the form

$$(au^2 + buv + cv^2)K(u, v) + wL(u, v, w) = 0, \tag{5.27}$$

where K, L are homogeneous polynomials, and the quadratic factor represents the two conjugate imaginary tangents $t, \bar{t}$ from O to $Q, \bar{Q}$. Since the line equation of O is $w = 0$, the other tangent lines from O to $\mathcal{A}$ are represented by

$$K(u, v) = 0. \tag{5.28}$$

We desire to vary the tangent lines $t, \bar{t}$ from O to $\mathcal{A}$, but not the others. This will be done if we vary the quadratic factor and keep $K(u, v)$, $L(u, v, w)$ the same; i. e., we define for $\mathcal{A}_\epsilon$:

$$[(a+a'\epsilon)u^2+(b+b'\epsilon)uv+(c+c'\epsilon)v^2]K(u,v)+wL(u,v,w)=0. \tag{5.29}$$

This gives explicitly the requisite variation of the algebraic curve $\mathcal{A}$.

We may now interpret for this general situation the directions α', α''; β', β''; ξ, η; previously considered for the normal form, with O located at the infinite point Y of the y-axis.

Let us consider one of the two tangents $t_\epsilon, \bar{t}_\epsilon$ from O to $\mathcal{A}_\epsilon$ defined by

$$(a + a'\epsilon)u^2 + (b + b'\epsilon)uv + (c + c'\epsilon)v^2 = 0, \tag{5.30}$$

for instance, t_ϵ. This intersects the fixed algebraic curve $\mathcal{A}$, regarded as a complete manifold $\mathfrak{F}$ of complex points (x, y), in two points $Q'_\epsilon, Q''_\epsilon$ near to Q, which come together in Q when $\epsilon = 0$. With the normal form of $\mathcal{A}$ previously considered, we saw that as ϵ passes through zero from positive to negative values, the points $Q'_\epsilon, Q''_\epsilon$ enter Q from opposite directions α', α'' and leave in the perpendicular opposite directions β', β'';[33] further, that the directions ξ, η of differentiation of Green's function are the angle-bisectors of $\alpha'\beta'$, $\alpha'\beta''$. By a previous italicized remark towards the beginning of this § 5. 2, all these angular relations continue to hold when we pass from the normal to the general form. Accordingly:

[33] The simplest example is the curve $y = x^2$ cut by the variable line $y = \epsilon$ (ϵ real). On the x-Riemann surface, the intersection points are $x = \pm\sqrt{\epsilon}$, which, as ϵ varies from positive values through zero to negative values, enter the origin from the opposite direction of the real axis and leave in the opposite directions of the imaginary axis.

In every case, the perpendicular ξ, η directions are precisely the angle-bisectors of the perpendicular α, β directions.

3. The variation just described depends on the real parameters a', b', c'. Of these, a' may be taken equal to zero, since any particular coefficient in (5. 30) may always be held fixed without loss of generality. Further, only the ratio $b' : c'$ is important, since any common factor of b', c' may be absorbed in ϵ. Thus, exactly one essential parameter is involved.[34] Evidently, this may be used to cause the ξ, η directions to coincide with *any preassigned pair of perpendicular directions through* Q.

4. We summarize in the following theorem, which constitutes the chief new result of the present paper. It seems of particular interest because of the interplay of fundamental geometric and analytic entities, as well as for its direct application to the problem of Plateau.

Variational Theorem Concerning Green's Function.

THEOREM. *Let $\mathcal{A}$ denote any real algebraic curve, on whose Riemann surface, $\mathcal{S}$ or $\mathcal{R}$, the points $Q, \bar{Q}$ are any two conjugate imaginary points.*

Let the tangent lines $t, \bar{t}$ to $\mathcal{A}$ at $Q, \bar{Q}$ intersect in the real point O. Choose a reference triangle with one vertex at O; then in homogeneous line coördinates u, v, w the equation of $\mathcal{A}$ will evidently have the form

$$\mathcal{A}: \quad (au^2 + buv + cv^2)K(u, v) + wL(u, v, w) = 0,$$

where

$$t, \bar{t}: \quad au^2 + buv + cv^2 = 0$$

represents the conjugate imaginary tangents $t, \bar{t}$. K and L are homogeneous polynomials.

Construct now the family of curves with real parameter ϵ,

$$\mathcal{A}_\epsilon: \quad [(a + a'\epsilon)u^2 + (b + b'\epsilon)uv + (c + c'\epsilon)v^2]K(u, v) + wL(u, v, w) = 0,$$

where a', b', c' are fixed but arbitrary real coefficients.

Then by rectilinear projection from O, the Riemann surfaces $\mathcal{S}$, $\mathcal{S}_\epsilon$ of $\mathcal{A}$, $\mathcal{A}_\epsilon$ are set into one-one, continuous, and conformal correspondence: P to $P(\epsilon)$, except in the immediate vicinity of $Q, \bar{Q}$. In particular, the real branches C, C_ϵ of $\mathcal{A}, \mathcal{A}_\epsilon$ are thereby set into one-one continuous correspondence. This depends on the circumstance that all the real tangents from O to $\mathcal{A}_\epsilon$, which are defined by the real factors of the equation $K(u, v) = 0$, remain invariant, since this equation is independent of ϵ.

---

[34] This may be taken as the ratio of any two chosen determinants in the matrix $\begin{vmatrix} a & b & c \\ a' & b' & c' \end{vmatrix}$.

In fact, for the same reason, all the tangents, real and imaginary, from O *to* $\mathcal{A}_\epsilon$ *except* $t_\epsilon, \bar{t}_\epsilon$ *remain fixed. These, however, vary with* ϵ*; and* t_ϵ*, for instance, intersects the Riemann surface* $\mathcal{S}$ *of* $\mathcal{A}$ *in two points* $Q'_\epsilon, Q''_\epsilon$ *near to* Q,[35] *which, as* ϵ *passes through the value zero from positive to negative, always enter* Q *from two opposite directions* α', α''*, and leave in the perpendicular opposite directions* β', β''*. Let the angle-bisectors of* $\alpha'\beta', \alpha'\beta''$ *be the perpendicular directions* ξ, η*.*

Denote by $G(P_1, P_2)$ *the Green's function of either conjugate semi-surface* R *of the symmetric Riemann surface* $\mathcal{S}$.[36] *This is for the case where* $\mathcal{S}$ *is orthosymmetric; if* $\mathcal{S}$ *is diasymmetric,* $G(P_1, P_2)$ *will denote the Green's function of* $\mathcal{S}'$*, which is* $\mathcal{S}$ *severed along its transition curves* C*, the real branches of* $\mathcal{A}$*.*

Then, under the precedingly described variation of the form of $\mathcal{A}$ *and the position of the points* P_1, P_2*, we have for Green's function the variational formula*

$$\delta G(P_1, P_2) = \frac{\partial G(P_1, Q)}{\partial \xi} \frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta} \frac{\partial G(P_2, Q)}{\partial \xi}$$

(apart from an inessential numerical factor).

Finally, the directions ξ, η *can be made to coincide with any preassigned perpendicular directions at* Q *by proper choice of* a', b', c'*.*

6. THE MOST GENERAL FORM OF THE PROBLEM OF PLATEAU.

There is no reason why, in the preceding theorem and proof, we should not allow $\mathcal{A}$ to be a general real *analytic* curve instead of merely an algebraic one. Using non-homogeneous line coördinates, with $v = -1$,[37] this amounts to permitting $K(u, -1) = K(u)$ and $L(u, -1, w) = L(u, w)$ to be *any convergent power-series* instead of restricting them to polynomials.

Hence the method of the present paper also solves the Plateau problem in what may be considered its most general form, where the required minimal surface M has a structure as to topology and boundaries of the generality of the Riemann surface of an arbitrary real analytic function.[38] That is, the

---

[35] The tangents $t, \bar{t}$ are supposed to have ordinary two-point contact with $\mathcal{A}$ at Q, Q. See the end of the first paragraph of § 5. 1.

[36] The notation is supposed arranged so that R contains Q, its conjugate R' containing $\bar{Q}$.

[37] The general line equation being then $y = ux + w$. The point O is the origin, whose line equation is $w = 0$.

[38] One whose power-series elements occur in conjugate complex pairs, i. e., with conjugate complex centers $x_0, \bar{x}_0$ and coefficients $a_m, \bar{a}_m$.

Riemannian manifold to which M is topologically equivalent may be an absolutely general one in the sense of Weyl.[39]

The paper immediately following this one in the present issue of this Journal is devoted to a detailed treatment of this extended form of the Plateau problem, *all the essential elements of whose solution are already contained in the present paper.* A note giving a precise formulation, a statement of results, and an indication of methods has already been published [12].

7. FORMULAS IN COMPLEX TERMS.[40]

It is a general fact that relations involving harmonic functions can also be expressed, often in quite an elegant manner, in terms of the complex analytic functions of which the harmonic ones are the real parts.

1. In the one-contour case of the problem of Plateau, a standard form of the Riemann surface R is the unit circle or the conformally equivalent half-plane $y > 0$.

For the half-plane we have

$$(7.1) \qquad G(P_1, P_2) = -\log \frac{\overline{P_1P_2}}{P_1\bar{P}_2} = -\tfrac{1}{2}\log \frac{(x_1-x_2)^2+(y_1-y_2)^2}{(x_1-x_2)^2+(y_1+y_2)^2}$$

where $\bar{P}$ is the reflection of P in the real axis C.

If y_1, y_2 are infinitesimal, then neglecting infinitesimals of higher order, we have

$$(7.2) \qquad G(P_1, P_2) = -\tfrac{1}{2}\log \frac{1+\left(\dfrac{y_1-y_2}{x_1-x_2}\right)^2}{1+\left(\dfrac{y_1+y_2}{x_1-x_2}\right)^2}$$

$$= -\tfrac{1}{2}\left\{\left(\frac{y_1-y_2}{x_1-x_2}\right)^2 - \left(\frac{y_1+y_2}{x_1-x_2}\right)^2\right\} = \frac{2y_1y_2}{(x_1-x_2)^2}.$$

Therefore

$$\lim_{y_1, y_2 \to 0} \frac{G(P_1, P_2)}{y_1y_2} = \frac{2}{(x_1-x_2)^2};$$

i. e., by (3.15),

$$(7.3) \qquad \frac{\partial^2 G(P_1, P_2)}{\partial n_1 \partial n_2} = \frac{2}{\overline{P_1P_2}^2}$$

---

[39] See H. Weyl, *Die Idee der Riemannschen Fläche*, 1913, p. 36; also L. Ahlfors, "Geometrie der Riemannschen Flächen," *Proceedings Oslo Congress*, vol. 1 (1936), pp. 239-248; S. Stoilow, "Sur la définition des surfaces de Riemann," *ibid.*, vol. 2, p. 143; S. Stoilow, *Principes Topologiques de la Théorie des Fonctions Analytiques*, 1938.

[40] The principal formulas of this section were first given by the author as follows: (7.6) in [2], also in the abstract immediately preceding [1], same journal, volume,

where P_1, P_2 are any two points on C, whose distance apart is $\overline{P_1P_2}$.

By substitution in (3.14), this gives for $A(g, R)$ the formula

$$A(g,R) = \frac{1}{4\pi}\int_C\int_C [g(P_1) - g(P_2)]^2 \frac{ds_1 ds_2}{\overline{P_1P_2}^2} \tag{7.4}$$

$$= \frac{1}{4\pi}\int_C\int_C [g(x_1) - g(x_2)]^2 \frac{dx_1 dx_2}{(x_1 - x_2)^2}.$$

We have

$$\frac{dx_1 dx_2}{(x_1 - x_2)^2} = \text{A. R. } (x_1, x_2, x_1 + dx_1, x_2 + dx_2)$$

where A. R. denotes "anharmonic ratio."

Let us regard the plane of R as a complex z-plane; then the real axis C is converted into the unit circumference by a transformation of the linear fractional form

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta}.$$

Since this leaves all anharmonic ratios invariant, it follows that if R now denote the unit circle and C its circumference, we have

$$A(g,R) = \frac{1}{4\pi}\int_C\int_C [g(z) - g(\zeta)]^2 \frac{dz d\zeta}{(z-\zeta)^2}, \tag{7.5}$$

where z corresponds to x_1, and ζ to x_2.

Let us introduce here the polar angles θ, ϕ by writing $z = e^{i\theta}$, $\zeta = e^{i\phi}$; then we easily get

$$A(g,R) = \frac{1}{4\pi}\int_C\int_C [g(\theta) - g(\phi)]^2 \frac{d\theta d\phi}{4\sin^2\dfrac{\theta-\phi}{2}}. \tag{7.6}$$

This was the author's original form $A(g)$ of the functional $A(g, R)$.

Evidently, then, the elementary kernel $dzd\zeta/(z-\zeta)^2$ of (7.5) is complex only in appearance, being equal to the positive real elementary kernel that occurs in (7.6).

2. Formulas (7.6), (7.5) illustrate two equivalent forms of $A(g, R)$, one in terms of real, the other in terms of complex quantities. This parallelism extends to the case where R is a general Riemann surface.

Let $G(P, Q)$ denote Green's function of R, while $z = x + iy$, $w = u + iv$ denote respectively the affixes of the points P, Q. Then $G(P, Q)$, as harmonic

---

and page; (7.5) in [3] and [6]; (7.14, 13, 10, 8) in [3]; (7.33, 36, 37) in [8], (7.39) in [9].

function of x, y, is always expressible as the real part of an analytic function of z in the form

$$G(P, Q) = -\Re\{S(z, w) - S(z, \bar{w})\}, \tag{7.7}$$

$\bar{w}$ being the symmetric point to w. This particular form of the complex function results from the fact that G is also harmonic in u, v, which implies $G = \phi(w) - \psi(\bar{w})$, and the symmetry of the complete Riemann surface $\mathfrak{R}$ serves to identify ϕ and ψ.

We have introduced also the functions

$$Z(z, w) = \frac{\partial S(z, w)}{\partial z} \qquad P(z, w) = \frac{\partial^2 S(z, w)}{\partial z \partial w}. \tag{7.8}$$

The notation S, Z, P is intended to recall the elliptic functions $\log \sigma, \zeta, \wp$ to which the former reduce in the two-contour case of the Plateau problem.

Let $H(u, v)$ denote the harmonic function on R with the boundary values $g(z)$ on C. Then

$$H(u, v) = \Re F(w). \tag{7.9}$$

A form of solution of the Dirichlet problem equivalent to (3.1) is to find $F(w)$ explicitly. We have given the formula

$$F(w) = \frac{1}{\pi i} \int_C g(z) Z(z, w)\, dz + \gamma, \tag{7.10}$$

where the constant γ is

$$\gamma = -\frac{1}{2\pi i} \int_C g(z) [Z(z, w_0) + Z(z, \bar{w}_0)]\, dz. \tag{7.11}$$

w_0 denotes any fixed point of R, whose assignment determines the additive pure imaginary constant inherent in $F(w)$, by the condition

$$\Im F(w_0) = 0. \tag{7.12}$$

Generally, $F(w)$ is not uniform on R, but $\Re F(w)$ always is; i. e., any periods of $F(w)$ are pure imaginary.

From (7.10) we easily derived

$$\sum_{i=1}^{n} F'_i{}^2(w) = \frac{1}{2\pi^2} \int_C \int_C [g(z) - g(\zeta)]^2 P(z, w) P(\zeta, w)\, dz d\zeta, \tag{7.13}$$

which serves the same purpose as the real formula (3.7).

Also, we obtained the following formula for $A(g, R)$ in complex terms:

$$A(g, R) = \frac{1}{4\pi} \int_C \int_C [g(z) - g(\zeta)]^2 P(z, \zeta)\, dz d\zeta. \tag{7.14}$$

This is the complex analogue of the real formula (3.14). Just as in the particular case (7.5), every element $P(z,\zeta)dzd\zeta$ is *positive real* in value, although its expression involves complex quantities.

3. For example, let R be a half-plane; then

$$G(P,Q) = -\Re \log \frac{z-w}{z-\bar{w}}, \tag{7.15}$$

evidently equivalent to (7.1).

By comparison with (7.7),

$$S(z,w) = \log (z-w). \tag{7.16}$$

Therefore, by (7.8),

$$Z(z,w) = \frac{1}{z-w}, \qquad P(z,w) = \frac{1}{(z-w)^2}. \tag{7.17}$$

The latter, substituted in (7.14), confirms (7.5).

In dealing with the variational condition, $\delta A(g,R) = 0$, in the one-contour case, the following easily verified algebraic identity was fundamental:

$$\frac{1}{(z-w)^2} + \frac{1}{(z-\zeta)^2} + \frac{1}{(\zeta-w)^2} = \left\{ \frac{1}{z-w} - \frac{1}{z-\zeta} - \frac{1}{\zeta-w} \right\}^2. \tag{7.18}$$

In terms of (7.17), this is expressible in the form

$$P(z,w) + P(z,\zeta) + P(\zeta,w) = \{Z(z,w) - Z(z,\zeta) - Z(\zeta,w)\}^2. \tag{7.19}$$

4. Let R denote the strip $0 \leqq \Re z \leqq \pi/2$. This is the appropriate Riemann surface for the Plateau problem when there are two contours with a single point in common.[41] We have here, as is easily verified,

$$G(z,w) = -\Re \log \frac{\sin (z-w)}{\sin (z-\bar{w})}, \tag{7.20}$$

where $\bar{w}$ is the *symmetric* (not conjugate complex) point to w, being the reflection of w in the imaginary axis, bounding R.

Here

$$S(z,w) = \log \sin (z-w), \qquad Z(z,w) = \cot (z-w), \qquad P(z,w) = \csc^2 (z-w). \tag{7.21}$$

We have the directly verifiable trigonometric identity:

[41] This is the case dealt with in [5].

$$(7.22)\qquad \csc^2(z-w)+\csc^2(z-\zeta)+\csc^2(\zeta-w) = \{\cot(z-w)-\cot(z-\zeta)-\cot(\zeta-w)\}^2.$$

In terms of Z and P, this is also expressed by the formula (7.19).

5. Let us consider the problem of Plateau for two contours. Then R is a circular ring, or semi-torus, and the complete Riemann surface $\mathfrak{R}$ is a torus. We have

$$(7.23)\qquad G(z,w) = -\Re\log\frac{\theta(v(z,w);\tau)}{\theta(v(z,\bar{w});\tau)}.$$

Here θ denotes the elliptic function of Jacobi (with characteristic 1/2, 1/2; $= -\theta_1(u;\tau)$), and $v(z,w)$ is the normal elliptic integral of the first kind on $\mathfrak{R}$ with the periods $1, \tau$, the latter being pure imaginary.

By comparison with (7.7, 8), we have

$$(7.24)\qquad S(z,w)=\log\theta(v(z,w)),\qquad Z(z,w)=\zeta(v(z,w))v'(z),$$
$$P(z,w)=\wp(v(z,w))v'(z)v'(w),$$

where

$$(7.25)\qquad \zeta(u)=\frac{d}{du}\log\theta(u),\qquad \wp(u)=-\frac{d^2\log\theta(u)}{du^2}.$$

These are not quite the elliptic functions ζ, $\wp$ of Weierstrass, but do not differ from them essentially. In fact,

$$(7.26)\qquad \sigma(u)=ae^{bu^2}\theta(u),$$

where a, b are constants as to u:

$$(7.27)\qquad a=\frac{1}{\theta'(0;\tau)},\qquad b=-\frac{\theta'(\frac{1}{2};\tau)}{\theta(\frac{1}{2};\tau)}.$$

σ denotes the elliptic function of Weierstrass, in whose notation,

$$(7.28)\qquad \zeta(u)=\frac{d}{du}\log\sigma(u),\qquad \wp(u)=-\frac{d^2\log\sigma(u)}{du^2}.$$

For the Weierstrass functions ζ, $\wp$ we have the classic addition theorems:

$$(7.29)\qquad \zeta(u_1+u_2)-\zeta(u_1)-\zeta(u_2)=\tfrac{1}{2}\frac{\wp'(u_1)-\wp'(u_2)}{\wp(u_1)-\wp(u_2)},$$

$$(7.30)\qquad \wp(u_1+u_2)+\wp(u_1)+\wp(u_2)=\tfrac{1}{4}\left\{\frac{\wp'(u_1)-\wp'(u_2)}{\wp(u_1)-\wp(u_2)}\right\}^2.$$

These combine to give

$$(7.31)\qquad \wp(u_1+u_2)+\wp(u_1)+\wp(u_2)=\{\zeta(u_1+u_2)-\zeta(u_1)-\zeta(u_2)\}^2.$$

If, instead, we use the ζ, $\wp$ based on θ according to (7.25), this relation

changes slightly by the acquisition of the constant term $-6b$ on the right. When we then write $u_1 = v(z, \zeta)$, $u_2 = v(\zeta, w)$, so that $u_1 + u_2 = v(z, w)$, we have [42]

$$(7.32) \qquad \begin{aligned} &\wp(v(z,w)) + \wp(v(z,\zeta)) + \wp(v(\zeta,w)) \\ &\quad = \{\zeta(v(z,w)) - \zeta(v(z,\zeta)) - \zeta(v(\zeta,w))\}^2 - 6b. \end{aligned}$$

Again, the analogy with (7.19) is apparent.

5. Let R denote a Riemann semi-surface with any finite number k of boundaries, but restricted to be of *genus zero*. The complete Riemann surface $\mathfrak{R}$ is of genus $p = k - 1$, according to (2.1).

Then we found

$$(7.33) \qquad G(z, w) = -\Re \log \frac{\theta(v(z,w)\,;\tau)}{\theta(v(z,\bar{w})\,;\tau)}.$$

Here, in generalization of (7.23), θ denotes the multiple θ-series of Riemann,[43] whose p arguments $v_j(z, w)$ are the normal integrals of the first kind on $\mathfrak{R}$. $\tau = |\tau_{jk}|$ $(j, k = 1, 2, \cdots, p)$ is the matrix of periods of v. The *characteristic* $\frac{1}{2}\beta$, $\frac{1}{2}\alpha$ of θ may consist of any $2p$ half-integers, such that $\beta\alpha = \beta_1\alpha_1 + \beta_2\alpha_2 + \cdots + \beta_p\alpha_p$ is an odd integer (odd characteristic).

Again comparing with (7.7, 8), we have

$$(7.34) \qquad \begin{aligned} S(z,w) &= \log\theta(v(z,w)), \qquad Z(z,w) = \zeta_j(v(z,w))v'_j(z), \\ P(z,w) &= \wp_{jk}(v(z,w))v'_j(z)v'_k(w), \end{aligned}$$

where

$$(7.35) \qquad \zeta_j(u) = \frac{\partial \log\theta(u)}{\partial u_j}, \qquad \wp_{jk}(u) = -\frac{\partial^2 \log\theta(u)}{\partial u_j \partial u_k}.$$ [44]

Concerning these functions, we proved the identity,[45] generalizing (7.32),

$$(7.36) \qquad \begin{aligned} &[\wp_{jk}(v(z,w)) + \wp_{jk}(v(z,\zeta)) + \wp_{jk}(v(\zeta,w))]v'_j(w)v'_k(w) \\ &\quad = [\zeta_j(v(z,w)) - \zeta_j(v(z,\zeta)) - \zeta_j(v(\zeta,w))] \\ &\qquad \times [\zeta_k(v(z,w)) - \zeta_k(v(z,\zeta)) - \zeta_k(v(\zeta,w))]v'_j(w)v'_k(w) + R(w). \end{aligned}$$

In this, $R(W)$ denotes an undetermined rational function on $\mathfrak{R}$ involving w alone.

We did not use this identity in its original form, but rather in the one obtained by applying the operator $\partial^2/\partial z\partial\zeta$ which, among other circumstances, eliminates the undetermined function $R(w)$. We obtained in this way:

---

[42] It should be easy to distinguish ζ the function from ζ the argument.

[43] See H. F. Baker, *Abel's Theorem and* $\cdots$ *θ-Functions*, 1897, chap. 10.

[44] A little later we shall also use $\wp_{jkl} = -\partial^3 \log\theta/\partial u_j\,\partial u_k\,\partial u_l$.

[45] In the next two formulas, the summation convention applies to all indices, which range from 1 to p, the genus of $\mathfrak{R}$.

$$(7.37)\quad \left\{ \begin{aligned} &\wp_{jkl}(v(z,\zeta))[-\zeta_m(v(z,w)) + \zeta_m(v(\zeta,w))] \\ &+ \wp_{kl}(v(z,\zeta))\wp_{jm}(v(z,w)) + \wp_{jm}(v(z,\zeta))\wp_{kl}(v(\zeta,w)) \\ &+ 2\pi i \frac{\partial \wp_{jk}(v(z,\zeta))}{\partial \tau_{lm}} \end{aligned} \right\} v'_j(z)v'_k(\zeta)v'_l(w)v'_m(w)$$
$$= \wp_{jl}(v(z,w))\wp_{km}(v(\zeta,w))v'_j(z)v'_k(\zeta)v'_l(w)v'_m(w).$$

The operator $dzd\zeta\partial^2/\partial z\partial\zeta$ is equivalent to $ds_1 ds_2 \partial/\partial n_1 \partial n_2$ as applied to (4.7) in order to get (4.11).

For elliptic functions, the identity (7.37) takes the following simpler form, where $u_1 = v(z,w)$, $u_2 = -v(\zeta,w)$, therefore $u_1 + u_2 = v(z,\zeta)$:

$$(7.37')\quad -\wp'(u_1+u_2)[\zeta(u_1)+\zeta(u_2)] + \wp(u_1+u_2)[\wp(u_1)+\wp(u_2)]$$
$$+ 2\pi i \frac{\partial \wp(u_1+u_2)}{\partial \tau} = \wp(u_1)\wp(u_2).$$

We emphasize again that here ζ, $\wp$ are derived from the Jacobi function $\theta_1(u;\tau)$ according to (7.25), not, as is more usual, from the Weierstrass function σ.

6. In the case of a general Riemann semi-surface R of *any finite genus* h and any finite number k of boundaries, we found, writing

$$(7.38)\qquad G(z,w) = -\Re\{S(z,w) - S(z,\bar{w})\},$$

that

$$(7.39)\quad S(z,w) = \log\theta(v(z,w);\tau) - T_{\alpha\beta}[v_\alpha(z) - v_{\alpha'}(z)][v_\beta(w) - v_{\beta'}(w)].$$

Here α, β are summation indices, corresponding to the h handles of R, and range from 1 to h, while $\alpha' = \alpha + h$, $\beta' = \beta + h$ range over the symmetric indices from $h+1$ to $2h$, corresponding to the h symmetric handles on the conjugate semi-surface R'. The quantities $T_{\alpha\beta}$ are certain functions of the periods τ, namely:

$$(7.40)\qquad T_{\alpha\beta} = \text{reciprocal matrix of } t_{\alpha\beta},$$

where

$$(7.41)\qquad t_{\alpha\beta} = \frac{1}{2\pi i}(\tau_{\alpha\beta} - \tau_{\alpha'\beta} - \tau_{\alpha\beta'} + \tau_{\alpha'\beta'});$$

i. e.,

$$(7.41')\qquad T_{\alpha\lambda}t_{\lambda\beta} = \begin{cases} 1 \text{ if } \alpha = \beta, \\ 0 \text{ if } \alpha \neq \beta. \end{cases}$$

The complementary terms in (7.39), as compared with (7.34), were easily taken account of, and led to an identity equivalent to (7.37).

6. All these identities can be summarized in the general formula:

$$(7.42)\qquad \delta S(z,\zeta) = Z(z,w)Z(\zeta,w)$$

when $\mathcal{R}$ or $\mathcal{A}$ is varied as explained in connection with the Variational Theorem of § 5.4. This is the complex form of the variational formula (1.3) for Green's function. (w denotes an arbitrary point Q of R.)

If we apply to (7.42) the operator $dzd\zeta\partial^2/\partial z\partial\zeta$, commutative with δ, we get, by means of (7.14, 13), the result

$$\delta A(g, R) = \frac{\pi}{2}\sum_{i=1}^{n} F'_i{}^2(w). \tag{7.43}$$

This is the complex form of (3.7).

7. We may summarize as follows.

The Variational Theorem concerning Green's function expresses a very general mathematical fact, which includes in rising stages the algebraic identity (7.18), *the trigonometric identity* (7.22), *the combined addition theorem* (7.31) *of the elliptic functions* $\zeta, \wp$, *and the identity* (7.36) *involving* θ*-functions on the Riemann surface of a general algebraic function.*

Since, according to § 6, *the symmetric Riemann surface* $\mathcal{R}$ *may even belong to any analytic function, the Variational Theorem may be regarded even more generally as an extension of the stated identities to what we may call hyper-*θ*-functions.*

8. ATTAINMENT OF THE MINIMUM OF $A(g, R)$.

1. It is now almost an anti-climax to complete our proof by establishing the *attainment* of the minimum of $A(g, R)$.

That this is a routine matter is indeed one of the advantages of our method. For our proof is isomorphic with the classic one of the attainment of the minimum of a lower semi-continuous function $f(x)$ of a real variable on a closed interval. This is, so to say, the "lower half" of the classic Weierstrass theorem where $f(x)$ is continuous and both maximum and minimum are attained.

As remarked by Fréchet,[46] the proof for this relatively elementary case applies to any real-valued function $U(a)$ on an abstract set $[a]$ of elements, provided that $[a]$ is an *L-set* which is *compact* and *closed*, and $U(a)$ is *lower semi-continuous*. An L-set means one where the notion of convergence or limit of a sequence of elements a_m has been defined in accordance with the following two postulates: (1) every sub-sequence of a sequence a_m which converges to a limit a^* is itself convergent to a^* as limit; (2) the "constant" sequence $a, a, a, \cdots$ converges to a. *Compact* means that every sequence a_m

[46] M. Fréchet, *Sur quelques points du calcul fonctionnel*, Palermo Rendiconti, v. 22 (1906), pp. 1-74.

contains a convergent sub-sequence. *Closed* means that the limit of any sequence of elements of $[a]$ is itself an element of $[a]$. *Lower semi-continuous* means: if $\lim a_m = a^*$, then $U(a^*) \leqq \liminf U(a_m)$.

The point is that if we include *improper* representations (g, R) of the given contours, as well as proper ones, then all these postulates are obeyed by our fundamental minimum problem

$$A(g, R) = \min. \tag{8.1}$$

The purpose of adjoining the improper representations is to produce closure of the set $[(g, R)]$. The compactness of $[(g, R)]$ follows from the fact that, first, R depends on only a finite number $3p - 3$ of parameters, and second, that the parametric representation g of the given contours is equivalent to a system of k monotonic functions.[47] It is well-known that any set of uniformly bounded monotonic functions is compact.[47a]

It is precisely the fact that the Fréchet postulates are here obeyed which makes the *immediate* solution of the minimum problem (8.1) safe from the criticism which Weierstrass applied to Riemann's use of the Dirichlet principle in the problem of the conformal mapping of plane regions. For in the Riemann-Dirichlet minimum problem:

$$\Omega(\omega) = \int\int [(\omega_x - \beta_y)^2 + (\omega_y + \beta_x)^2] dxdy = \min.,$$

the argument range $[\omega]$ is not compact. It was not till about a half-century after Riemann's work that Hilbert succeeded in proving rigorously the attainment of the minimum of $\Omega(\omega)$.[48]

2. We list the improper representations (g, R) as follows.

(I) R improper.

(a) R *reduces*; i. e., separates into two disconnected parts, or has its topological characteristic lowered to less than the prescribed value r. Both these phenomena are due to coalescence of branch-points of R.

(b) Two boundary curves C of R, or real branches of the corresponding algebraic curve $\mathcal{A}$, touch in a point, whereas the corresponding contours Γ

[47] Cf. [2], § 3; [9], § 13.

[47a] In analytic form due to Helly, in geometric form (with an equivalent condition of uniformly bounded length) to Fréchet, *loc. cit.* See also Hilbert, *Jahresbericht der Deutschen Mathematiker-Vereinigung*, vol. 8 (1899), p. 184.

[48] Cf., for this paragraph, the author's paper "Remarks on Riemann's doctoral dissertation," *Proceedings of the National Academy of Sciences*, vol. 24 (1938), pp. 297-302.

have, as we assume, no point in common. That is, the curve $\mathcal{A}$ has a real double- (or multiple-)point.

(II) g improper.

(a) *Discontinuously improper,* where a *partial* arc of a contour Γ corresponds to a single point P of a real branch C.

(b) *Continuously improper,* where a *partial* arc of a real branch C corresponds to a single point α of a contour Γ.

(c) *Degenerate,* where an *entire contour* Γ corresponds to a single point P of a real branch C, and the *entire real branch* C corresponds to a single point α of Γ.

All these improper forms are limits of proper ones, as is easily seen.

3. We must next extend the definition of $A(g, R)$ to all improper representations. Only the cases (Ia) and (IIc) require discussion.

If R separates:

$$R = R_1 + R_2 + \cdots + R_m, \tag{8.2}$$

we define, quite naturally,

$$A(g, R) = A(g_1, R_1) + A(g_2, R_2) + \cdots + A(g_m, R_m), \tag{8.3}$$

where $g_1, g_2, \cdots, g_m$ denote the parts of g that go with the boundaries of $R_1, R_2, \cdots, R_m$ respectively.

Let g degenerate completely, as described in (IIc), with respect to the contour Γ_1. Then we have found [49] that the definition

$$A(g, R) = d(\Gamma_1) + A(\alpha, g_2, \cdots, g_k; R) \tag{8.4}$$

is uniquely determined by the lower semi-continuity property in the exact form: [50]

$$A(g, R) = \min \lim A(g_1, \bar{g}_2, \cdots, \bar{g}_k; \bar{R}), \tag{8.5}$$

where g_1 tends to degeneration while always representing Γ_1 parametrically, and $\bar{g}_2, \cdots, \bar{g}_k; \bar{R}$ tend to $g_2, \cdots, g_k; R$. In this definition,

$$d(\Gamma_1) = \min A(g_1) \tag{8.6}$$

for all parametric representations g_1 of the contour Γ_1; the concrete geometric interpretation of $d(\Gamma_1)$ is the area of the simply-connected minimal surface bounded by Γ_1.

---

[49] [9], § 14.4. For the two-contour case, cf. [3], p. 342.

[50] That is, with equality of the functional value to the inferior limit, instead of merely $\leqq$.

Similarly, if the parametric representations of several contours degenerate, e. g., $g_1, g_2, \cdots, g_j$ into the points $\alpha_1, \alpha_2, \cdots, \alpha_j$, then the unique definition which accords with (8. 5) is

$$A(g, R) = d(\Gamma_1) + d(\Gamma_2) + \cdots + d(\Gamma_j) + A(\alpha_1, \alpha_2, \cdots, \alpha_j, g_{j+1}, \cdots, g_k; R). \tag{8.7}$$

In the most fundamental case of one contour, the degenerate representations can be completely avoided by means of a three-point condition, based on the fact that $A(g)$ is invariant under linear fractional transformation of z,

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta}, \tag{8.8}$$

with three parameters $\alpha : \beta : \gamma : \delta$.[51]

5. $A(g, R)$ *is lower semi-continuous,* because the kernel $\partial^2 G(P_1, P_2)/\partial n_1 \partial n_2$ of its expression (3. 14) as an integral is *positive.*[52] We have defined

$$A(g, R) = \lim_{\epsilon \to 0} A_\epsilon(g, R), \tag{8.9}$$

where $A_\epsilon(g, R)$ is defined by the same integral as $A(g, R)$, but with the variables of integration restricted by the condition $\overline{P_1 P_2} \geqq \epsilon$. This prevents the integrand from becoming infinite, as would otherwise happen, due to the singularity $2/\overline{P_1 P_2}^2$ when P_1 and P_2 approach to coincidence.

Since we may pass to the limit under an integral sign provided that the integrand stays uniformly bounded, it follows that $A_\epsilon(g, R)$ is a continuous functional of (g, R). All the elements of $A(g, R)$ are positive; therefore $A(g, R)$ is the *increasing* limit of $A_\epsilon(g, R)$. Now, it is an easily proved theorem that the increasing limit of a continuous functional is lower semi-continuous.[53]

6. Therefore the minimum value of $A(g, R)$ is certainly attained for some particular representation (g^*, R^*), proper or improper.

All that remains is to exclude the possibility that (g^*, R^*) be improper.

In the one-contour case the only condition required is

$$d(\Gamma) < \infty, \tag{8.10}$$

i. e., finite. For here the Riemann surface R does not vary, but is constantly the half-plane or unit circle. Therefore we have only to avoid the singu-

[51] See the remarks just preceding (7. 5). α, β, γ, δ are real if we take R to be the half-plane $y > 0$.

[52] See the remarks associated with (3. 16).

[53] Cf. [2], p. 282.

larities (II). (IIc) is excluded by means of a three-point condition, as just before explained. (IIb) is avoided by means of the variational condition $\sum_{i=1}^{n} F'_i{}^2(w) = 0$, with which it is inconsistent.[54] For (IIa) we require the hypothesis (8.10). In fact, for all discontinuously improper representations,

$$A(g, R) = +\infty. \tag{8.11}$$

To see this, let α, β be the necessarily distinct[55] end-points of the partial arc of Γ which corresponds to the point P, and let $l^2 = (\text{dist } \alpha\beta)^2 = \sum_{i=1}^{n} (\alpha_i - \beta_i)^2$; then $l > 0$. If P_1, P_2 approach to P along C from *opposite* sides, then the integrand of $A(g, R)$ is asymptotically

$$\frac{2l^2}{\overline{P_1P_2}^2}. \tag{8.12}$$

The order to which this becomes infinite for $P_1 \to P$, $P_2 \to P$ leads easily to (8.11).

In the general case, however, R is variable, and we have to exclude the possibilities (Ia), (Ib) as well as the degenerate form (IIc). To the last mentioned the three-point condition no longer applies. For it is only when $p = 0$ that the complete Riemann surface $\mathfrak{R}$ of genus p has a three-parameter group of conformal automorphisms;[56] when $p = 1$, the number of parameters of the group is only one, and when $p > 1$, there is in general no conformal automorphism of $\mathfrak{R}$ except the identity.

7. Sufficient conditions for the solution. It is in order to exclude the possibility of reduction of R according to (Ia) that we need additional sufficient conditions.[57]

We consider the following three *lower bounds*:

$$d(\Gamma, r\ ) = \min A(g, R\ ), \tag{8.13}$$

$$d(\Gamma, r'\ ) = \min A(g, R'\ ), \tag{8.14}$$

$$d(\Gamma, r'') = \min A(g, R''). \tag{8.15}$$

Here R denotes any *proper* Riemann surface with k boundaries and of topological type r or $\bar{r}$ (the latter, we recall, indicating the one-sided case).

[54] See the last article of this section.

[55] For Γ is a Jordan curve—has no double-points.

[56] Which respect the symmetry of the surface.

[57] [9], § 5.

R' denotes all *primary reduced* forms of R. R'' includes all reduced forms of R, primary or composite.

Then we can easily show that [58]

$$d(\Gamma, r) \leqq d(\Gamma, r'), \tag{8.16}$$

$$d(\Gamma, r') = d(\Gamma, r''). \tag{8.17}$$

These relations result directly from the consideration that any improper representation (g, R') (or (g, R'')) can be modified slightly so as to form a proper one with only slight change in the value of $A(g, R)$. Also, any composite reduced form of representation (g, R'') can be modified slightly so as to give a primary reduced form (g, R') with only slight change of the value of $A(g, R)$. This gives the relation $\leqq$ in (8.17); the reverse inequality $\geqq$ follows from our interpretation of the type r'' as including r'.

The equality (8.17) justifies considering among the reduced forms R', R'' only the primary reduced type R'. By reference to § 2.6, we see that $d(\Gamma, r')$ is the least of the quantities

$$d(\Gamma, r-2), \qquad d(\Gamma_1, r_1) + d(\Gamma_2, r_2), \tag{8.18}$$

formed for all partitions

$$\Gamma = \Gamma_1 + \Gamma_2, \qquad r = r_1 + r_2 \tag{8.19}$$

of the ensemble of given contours and the topological characteristic r.

In the one-sided case, $d(\Gamma, \bar{r}')$ is the least of

$$\begin{aligned} &d(\Gamma, \overline{r-1}), \qquad d(\Gamma, r-1), \\ &d(\Gamma_1, \bar{r}_1) + d(\Gamma_2, \bar{r}_2), \qquad d(\Gamma, \bar{r}_1) + d(\Gamma_2, r_2). \end{aligned} \tag{8.20}$$

It is understood that only even values of r, r_1, r_2 can apply to any two-sided surface, but for a one-sided surface these characteristics may be odd or even.

The concrete geometric interpretation of $d(\Gamma, r)$ is

$$d(\Gamma, r) = a(\Gamma, r) = \min \mathfrak{A}(S_r), \tag{8.21}$$

where $\mathfrak{A}(S_r)$ denotes the area of S_r, which ranges over all surfaces of type r bounded by the contours Γ. But, as is apparent, our analysis is entirely independent of this fact, being based exclusively on the functional $A(g, R) = D(H)$, rather than on the area functional $\mathfrak{A}(S)$.

Besides

$$d(\Gamma, r) < +\infty, \tag{8.22}$$

[58] [9], §§ 5, 16, 19.

let us now make the additional hypothesis that actually

$$d(\Gamma, r) < d(\Gamma, r'), \tag{8.23}$$

i.e., that the strict inequality holds in (8.16).

Then in the minimizing representation *the Riemann surface R cannot have the reduced form R' or R''. For*

$$A(g, R'') \geqq d(\Gamma, r'') = d(\Gamma, r') > d(\Gamma, r) \tag{8.24}$$

which contradicts the condition of attainment of the minimum:

$$A(g, R'') = d(\Gamma, r). \tag{8.25}$$

8. The degenerate representations (IIc) are eliminated in quite the same way. For it follows easily from (8.7) that if g_0 is degenerate,

$$A(g_0, R) \geqq d(\Gamma, r'); \tag{8.26}$$

this is because the second member of (8.7) represents a particular mode of degeneration of R, with the detachment of simply-connected parts.

9. The discontinuously improper representations (IIa) are eliminated by the hypothesis (8.22) and the relation (8.11).

The same reasons eliminate the improper form (Ib). For let the real branches C_1, C_2 touch at the point P, which must correspond to *distinct* points α_1 of Γ_1 and α_2 of Γ_2. Then, for P_1 on C_1 and P_2 on C_2 both near to P, the integrand of $A(g, R)$ again has the asymptotic form (8.12).

10. Finally, we exclude the continuously improper representations (IIb).[59] This is easily done by means of the variational condition $\delta A(g, R) = 0$, leading to $E = G$, $F = 0$, which expresses the conformal nature of the correspondence between R and the minimal surface M. For if an arc of C is converted into a single point of Γ, then the modulus of conformality λ is zero all along this arc. On account of the connection of λ with harmonic or complex analytic functions, it follows readily that λ is identically zero. That is, the conformal transformation of R into M must convert all of R into a single point. But this is absurd, since the given contours Γ certainly do not reduce to a single point.

9. ARBITRARY JORDAN CONTOURS.

By a simple limit process, we can lift the restriction (8.22) of finite $d(\Gamma, r)$. This condition is equivalent to the supposition that each contour is capable of bounding a simply-connected surface of finite area.[60]

---

[59] Cf. [2], § 18; [9], § 15.4.

[60] A sufficient (but not necessary) condition for this is the rectifiability of each contour.

Let the contours Γ then be perfectly general Jordan curves, where usually

$$d(\Gamma, r) = +\infty. \tag{9.1}$$

In the case $d(\Gamma, r)$ finite, we define

$$e(\Gamma, r) = d(\Gamma, r') - d(\Gamma, r) \geqq 0, \tag{9.2}$$

by (8.16).

Whether $d(\Gamma, r)$ is finite or infinite, we then define

$$\bar{e}(\Gamma, r) = \max \overline{\lim_{m\to\infty}}\, e(\Gamma_m, r) \geqq 0, \tag{9.3}$$

where Γ_m denotes any sequence of contour systems tending to Γ such that[61] $d(\Gamma_m{}^j)$ is finite for every contour Γ^j of every system Γ_m. For instance, all the approaching contours may be polygons. $\overline{\lim}$ denotes the superior limit. The maximum is with respect to all possible sequences Γ_m.

We have proved in [9], § 17 the following theorem:[62]

If in (9.3) $\bar{e}$ *is strictly positive:*

$$\bar{e}(\Gamma, r) > 0, \tag{9.4}$$

then the contours Γ *are the boundaries of a minimal surface of topological type* r.

This minimal surface M is such in the sense of the Weierstrass formulas (1.11, 12); in the general case, M has infinite area, like every other surface bounded by the contours Γ.

The purpose of the condition (9.4) is to obviate the possibility of reduction of the Riemann surface R, conformally equivalent to M.

The preceding discussion applies throughout to one-sided surfaces with $\bar{r}$ in place of r.

10. CONCLUDING REMARKS.

The foregoing developments would seem to be only the beginning of a general theory of calculus of variations problems where extremal surfaces are sought of any prescribed topological form and with any number of assigned boundaries.

Sufficient conditions like

$$d(\Gamma, r) < d(\Gamma, r')$$

[61] In the notation $d(\Gamma, r)$, we omit r when it is zero.

[62] Announced in [8], p. 108.

may be invoked in order to prevent the degeneration of the extremal surface. Of course, here $d(\Gamma, r)$ represents the lower bound of whatever integral is to be minimized.

One-sided surfaces can be included quite as well as two-sided by the standard device of a two-sided covering surface.

BROOKLYN, N. Y.

REFERENCES

The following is a list of those publications by the author which are referred to in the course of the preceding exposition. Of these, the most comprehensive is [9], which includes a review of the essential features of all the others. There, a more complete bibliography is given, which refers also to the important work on the Plateau problem of T. Radó, R. Courant, and others.

1. "A general formulation of the problem of Plateau," *Bulletin of the American Mathematical Society*, vol. 36 (1930), p. 50. Abstract of a paper presented to the Society, October, 1929.

2. "Solution of the problem of Plateau," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 263-321.

3. "The problem of Plateau for two contours," *Journal of Mathematics and Physics*, vol. 10 (1931), pp. 310-359.

4. "One-sided minimal surfaces with a given boundary," *Transactions of the American Mathematical Society*, vol. 34 (1932), pp. 731-756.

5. "Crescent-shaped minimal surfaces," *Proceedings of the National Academy of Sciences*, vol. 19 (1933), pp. 192-199.

6. "The problem of Plateau," *Bulletin of the American Mathematical Society*, vol. 39 (1933), pp. 227-251.

7. "Some new results in the problem of Plateau," *Journal of Mathematics and Physics*, vol. 15 (1936), pp. 55-64.

8. "Minimal surfaces of general topological structure with any finite number of assigned boundaries," *ibid.*, vol. 15 (1936), pp. 105-123.

9. "Minimal surfaces of higher topological structure," *Annals of Mathematics*, vol. 40 (1939), pp. 205-298.

10. Same title as [9], being a preliminary note in the *Proceedings of the National Academy of Sciences*, vol. 24 (August, 1938), pp. 343-353.

11. Same title as the present paper, being a preliminary note. *Ibid.*, pp. 353-360.

12. "The most general form of the problem of Plateau," *ibid.*, pp. 360-364, a preliminary note to the paper which follows.

THE MOST GENERAL FORM OF THE PROBLEM OF PLATEAU.*

By JESSE DOUGLAS.

In a note published in the Proceedings of the National Academy of Sciences,[1] bearing the same title as the present paper, the author formulated the problem of Plateau in what may be considered its most general form, and gave the essential features of its solution, including a statement of results in the form of two theorems. This note immediately followed one which summarized the preceding paper in the present volume of this Journal, and had the same title: "Green's function and the problem of Plateau." As remarked in § 6 of "GF,"[2] it contains essentially also the proofs of the theorems stated in the first mentioned note for the most general case of the problem. However, this general case seems of sufficient importance to deserve a separate paper by itself—this is the one at hand.[3]

1. Formulation. We repeat the formulation of the general problem as given in the cited note.

PROBLEM P. *Given any riemannian manifold R in the most general sense of the term, i. e., any two-dimensional topological variety for which there is defined in the neighborhood of each point a local conformal representation on a circle.*[4] *R may then have any finite or infinite number of boundaries, and any topological structure whatever—i. e., any finite or infinite type of connectivity. It may also have either character of orientability, i. e., one- or two-sidedness.*

Given also any point-set Γ in n-dimensional euclidean space which is a topological image of the total boundary C of R. Γ may consist of any finite or infinite number of Jordan curves, together with their limit points; or it may be some more general type of point-set. A definite sense of description

* Received August 15, 1938.

[1] [12] of the list at the end of the foregoing paper.

[2] In this way we shall refer to the foregoing paper.

[3] To avoid too many back references, we shall repeat the necessary parts of "GF" in paraphrased form at the appropriate places.

[4] Subject to certain postulates, which express essentially that this conformal representation is self-consistent whenever two neighborhoods overlap. See Weyl, *Die Idee der Riemannschen Fläche*, 1913, p. 36; T. Radó, "Über den Begriff der Riemannschen Fläche," *Acta Litt. ac Scient. Szeged*, vol. 2 (1925), pp. 101-121; S. Stoilow, *Principes Topologiques de la Théorie des Fonctions Analytiques*, 1938.

is associated with each component of Γ, *carried over from* C, *which may be supposed oriented so that* R *is on the left.*

To determine the existence of a minimal surface M *topologically equivalent to* R *and bounded by* Γ.

2. Riemann surfaces. If the boundary of R consists of a finite number k of closed curves, and its topological characteristic r, or genus h, is finite, then we have the most important special case of Problem P, which we shall call Problem P_0—this is the one solved in the preceding paper. In Problem P_0, R is the Riemann semi-surface of a real *algebraic* curve; in the general problem, as will be brought out in the next section, R belongs similarly to a real [5] *analytic* curve $\mathcal{A}$, or function $y = f(x)$.

The complete Riemann surface $\mathcal{R}$ of $y = f(x)$ is *symmetric*, i. e., its points occur in conjugate complex pairs, $P, \bar{P}$. The *Riemann semi-surface* R is the manifold whose individual elements are the conjugate pairs $P, \bar{P}$, regarded as a single geometric entity. The boundaries of R correspond to the *real branches* C of the analytic curve $\mathcal{A}$. If these separate the complete surface $\mathcal{R}$, then this is called *orthosymmetric*, and the semi-surface R is a two-side or orientable manifold. If the real branches C do not separate $\mathcal{R}$, but rather we can pass on $\mathcal{R}$ from any point P to its conjugate $\bar{P}$ without traversing C, then $\mathcal{R}$ is called *diasymmetric*, and R is a one-sided or non-orientable manifold.

If all the points (x, y) of the curve $y = f(x)$ are plotted in the four-dimensional space of the two complex variables $x = x_1 + ix_2$, $y = y_1 + iy_2$, or of the four real variables x_1, x_2, y_1, y_2, the resulting two-dimensional analytic manifold—which we shall denote by $\mathcal{S}$—may be taken as one of the forms of the Riemann surface of $y = f(x)$. The real branches of $\mathcal{A}$ are formed by the intersection of $\mathcal{S}$ with the real x_1y_1-plane. The orthogonal projection of $\mathcal{S}$ on the complex x-plane ($x = x_1 + ix_2$) is the Riemann surface $\mathcal{R} = \mathcal{R}_x$ of $\mathcal{A}$ in its standard many-sheeted form, with any finite or denumerably infinite number of branch-points of finite or infinite order. Similarly, we have the Riemann surface $\mathcal{R}_y$ of the inverse function $x = f^{-1}(y)$, which is the orthogonal projection of $\mathcal{S}$ on the complex plane of $y = y_1 + iy_2$. These three forms $\mathcal{S}$, $\mathcal{R} = \mathcal{R}_x$ and $\mathcal{R}_y$ of the Riemann surface of $\mathcal{A}$ are all in conformal correspondence by the orthogonal projection, and therefore equivalent for all our purposes. Two analytic curves $\mathcal{A}, \mathcal{A}_1$ whose Riemann surfaces are conformally equivalent will be said to be related by a *biuniform transformation*. If the two curves are algebraic, we have a *birational* transformation. In the case where the curves are real, we require the biuniform transformation

[5] See footnote 38 of "GF."

to respect symmetric or conjugate points; then this transformation is also called *real.* Any two real biuniformly equivalent analytic curves are completely interchangeable as far as the purposes of the present paper are concerned.

3. Riemannian manifold, real analytic curve, normal polygon. It is a well-known fact, going back to Riemann himself, that every *closed* riemannian manifold[6] is exactly the Riemann surface of an *algebraic* function or curve, which is determined to within arbitrary birational transformation.

Much later (1909), this theorem was extended by Koebe to the most general riemannian manifold R with a proof that any given R is always exactly the Riemann surface of some *analytic* function or curve.[7] Subsequently Koebe gave this theorem in the form that an arbitrary orientable riemannian manifold R can always be represented as the Riemann semi-surface of an *orthosymmetric, real* analytic curve $\mathcal{A}$.[8] This result is easily adapted also to non-orientable manifolds R, which are thus shown to be always equivalent to the Riemann semi-surface of a diasymmetric real analytic curve $\mathcal{A}$. Further, the analytic curve $\mathcal{A}$ is always "complete" in the sense defined by Koebe; i. e., its Riemann surface has only "improper" openings.[9]

We shall assume this result of Koebe's, and hence *take the riemannian manifold R to be given from the start as the Riemann semi-surface of a real analytic curve $\mathcal{A}$, whose real branches C then constitute the boundary of R.*

We shall take it as a classical result, contained in the rigorous work of Hilbert and his followers on the Dirichlet principle, that Green's function $G(P, Q)$ always exists for R.[10] Also, we shall suppose as known that the Dirichlet problem is always solvable for R, in other words, that a harmonic function $H(Q)$ exists on R with any preassigned continuous[11] values $g(P)$ on the boundary C. Furthermore, we shall assume that $H(Q)$ is given

---

[6] The closure of the manifold implies a finite genus.

[7] P. Koebe, *Comptes Rendus*, Paris, v. 148 (1909), p. 1446. See the detailed presentation of an essentially equivalent proof in the cited book by Stoilow, pp. 43-62, at the end of which additional references will be found.

[8] P. Koebe, "Allgemeine Theorie der Riemannschen Mannigfaltigkeiten," *Acta Mathematica*, vol. 50 (1927), pp. 29-157; see p. 152.

[9] With "complete" we translate Koebe's term "vollkommen." An "improper" (or parabolic) opening ("uneigentliche Öffnung") is one conformally equivalent to that produced by removing a single point, e. g., the center, from a circular disc. In contrast, an "eigentliche" or "hyperbolische Öffnung" is conformally equivalent to the region adjacent to a circumference on its exterior side.

[10] Hilbert, *Mathematische Annalen*, vol. 59 (1904), pp. 161-186; Weyl, *loc. cit.*, p. 93 et seq.

[11] Certain types of discontinuity may also be allowed, as required for the improper representations g discussed in § 9.

explicitly by the same classic formula that applies in the case of a finite number of boundaries and finite connectivity of R, namely:

$$H(Q) = \frac{1}{2\pi}\int_C g(P)\,\frac{\partial G(P,Q)}{\partial n}\,ds. \tag{1}$$

Another standard representation of a general riemannian manifold R is as *normal polygon* $\mathfrak{P}$ in the non-euclidean plane of Lobatschevsky.[12] The complete N. E.[13] plane may be realized in the Poincaré model as interior of the unit circle, the N. E. straight lines being represented by circular arcs orthogonal to the unit circumference. The normal polygon $\mathfrak{P}$ is the fundamental region of a properly discontinuous group γ of N. E. motions, where γ contains no rotations, i. e., has no fixed point in the finite part of the N. E. plane. The N. E. motions γ are representable in the form

$$t' = \frac{\alpha_i t + \beta_i}{\gamma_i t + \delta_i} \qquad (i = 0, 1, 2, 3, \cdots) \tag{2}$$

of linear fractional transformations of the unit circle into itself. The fundamental region $\mathfrak{P}$ of γ is a polygon bounded by a finite or infinite number of N. E. straight lines—i. e., arcs of circles orthogonal to the unit circumference—together with a finite or infinite number of arcs of this circumference.[14] The former arcs correspond point-wise in pairs by transformations of γ; it is the identification of such corresponding points on its boundary which enables $\mathfrak{P}$ to represent a manifold R of arbitrary topological structure. The part of the boundary of $\mathfrak{P}$ which lies upon the unit circumference represents the boundary of R.

4. Examples. Generally, we may use k, finite or infinite, to represent the number of closed curves which form the given boundary Γ or its homeomorph C, and indicate the topological form of R or the required minimal surface M either by h, the genus, or r, the characteristic, together with the one- or two-sided nature. h may be interpreted concretely as the number of "handles" of the surface; r is the maximum number of circuits no linear combination of which separates the surface, also the excess of the connectivity over the number of boundaries. In the "finite"[15] or "algebraic" case, $r = 2h$ when the surface is two-sided; when the surface is one-sided, r may be any integer $\geqq 1$, odd or even.

---

[12] See Weyl, *loc. cit.*, §§ 19, 20.

[13] Throughout, "N. E." is an abbreviation for "non-euclidean."

[14] Or more general point-set—in any case, a "fundamental region" of the group γ regarded as acting upon the unit circumference.

[15] k finite, r finite.

In the finite case, the two integers k and h, or k and r, together with the character of orientability, characterize the surface topologically with completeness, but in the "infinite" or "analytic" case it no longer suffices for this purpose merely to give the value of k or h or both as infinity. For example, we may construct in a spherical surface an infinite number of perforations, distributed in two different ways, so that the resulting surfaces are topologically non-equivalent—for instance, one set of perforations may have a unique limit point, the other, two or many such points. Similarly, an infinite number of handles, growing smaller and smaller, attached to the spherical surface, will produce topologically distinct manifolds if distributed first so as to have a unique limit point, then so as to have many such points. In the case where the surface is one-sided, with an infinite number of branch-cuts or handles, these may likewise be distributed in topologically distinct ways.[16]

The following examples illustrate various particular forms, as to topological structure and boundaries, of the riemannian manifold R, or its homeomorph, the required minimal surface M.

Two-sided case.

(i) $y = \sqrt{\sin x}$. Here $k = \infty$, $h = 0$. R is equivalent to a sphere with an infinite number of perforations clustering towards a single limit point.

(ii) $y = \sqrt{\sin x}\,\sqrt{P(\sqrt{x})}$, where P denotes a polynomial of degree $4h + 4$ with real coefficients and one whose roots in $\sqrt{x}$ are neither real nor pure imaginary, so that the corresponding values of x, over which lie branch-points of R, are imaginary. Then $k = \infty$, h is finite; R is equivalent to a sphere with h handles and an infinite number of perforations having a unique limit point.

(iii) $y = \sqrt{\sinh \sqrt{x}}$. $k = 1$, $h = \infty$. The handles approach to a unique limit point.

(iv) $y = \sqrt{\sin x \sinh \sqrt{x}}$. $k = \infty$, $h = \infty$, with unique limit point for the perforation and for the handles.

One-sided case.

(v) $y = \sqrt{(x^2 + 1)\sin x}$. The form of R or M is that of a Möbius strip in which have been made an infinite number of perforations converging to a unique limit point.

---

[16] Exact conditions necessary and sufficient for the homoemorphism of any two surfaces of the most general possible topological structure have been given by B. v. Kerékjártó; see his *Vorlesungen über Topologie*, 1923, p. 170, also Stoilow, *loc. cit.*, pp. 93-101.

(vi) $y = \sqrt{P(x)\cosh x}$, where $P(x)$ is a polynomial with $2k$ distinct real roots. The surface is one-sided, with k boundaries, and an infinite number of branch-cuts converging to a unique limit point.

(vii) $y = \sqrt{\sin x \cosh x}$. A one-sided surface with an infinite number of boundaries and an infinite number of branch-cuts, with a unique limit point both for the boundaries and for the branch-cuts.

5. Exact formulation of the general Plateau problem. By a minimal surface M, we shall understand a *harmonic* and *conformal* image of any riemannian manifold R. The formulas which express this are [17]

$$\mathfrak{x} = H(Q), \tag{3}$$

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0. \tag{4}$$

Here $\mathfrak{x}$ denotes a vector locating an arbitrary point in n-dimensional euclidean space. $H(Q)$ denotes a vector whose n components are *harmonic functions* of the argument Q, an arbitrary point of R. ξ, η are *any two perpendicular directions* on R at the point Q. The dot denotes the scalar product of two vectors.

According to the Dirichlet principle, the harmonic function $H(Q)$ is completely determined by its values $g(P)$ on the boundary C of R. The boundary of the minimal surface M then consists of the locus of the equation $\mathfrak{x} = g(P)$.[18] The entire point of the Plateau problem is to make this coincide with a *preassigned* locus Γ, generally consisting of a finite or infinite number of contours (Jordan curves) together with their limit points.

We are now in a position to formulate in precise analytic terms the most general form of the problem of Plateau, stated at the beginning as Problem P.

Given an arbitrary riemannian manifold R, in the form of the Riemann semi-surface of a real analytic curve $\mathcal{A}$, and any locus Γ in n-dimensional euclidean $\mathfrak{x}$-space which is a topological image of the total boundary C of R, i.e., of the real branches of $\mathcal{A}$.

To find in the class $\mathfrak{J}$ of riemannian manifolds topologically equivalent to R a particular one R^*, and upon R^* a harmonic vector $H(Q)$ remaining one-valued and continuous on the boundary C^*, such that

(α) *the boundary values of $H(Q)$ on C^*,*

$$\mathfrak{x} = g^*(P),$$

form a parametric representation of Γ;

---

[17] Compare "GF," formulas (1.8), (1.14).

[18] Like $\mathfrak{x}$ and H, g represents a *vector*.

(β) *for every point Q on R^* and every pair of perpendicular directions ξ, η at Q,*

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0.$$

Then, evidently, the equation

$$\mathfrak{x} = H(Q)$$

represents a minimal surface M of the topological type $\mathfrak{T}$ of R and bounded by Γ.

In the case where R is one-sided, we require $H(Q)$ to be defined on the complete diasymmetric Riemann surface $\mathfrak{R}$ whose semi-surface is R, with the additional stipulation:

(γ) $$H(Q) = H(\bar{Q}),$$

where $Q, \bar{Q}$ denote any pair of symmetric points of $\mathfrak{R}$.

$H(Q)$ is not required to remain regular on the boundary curves C of R, but only one-valued and continuous. In other words, in solving the Dirichlet problem for $\mathfrak{R}$ in the diasymmetric case, with continuously varying assigned values $g(P)$ on the real branches C, we may regard $\mathfrak{R}$ as severed along C. Then the curves C become boundaries for the severed manifold $\mathfrak{R}'$, and at any points P', P'' of these boundaries which were originally coincident in a point P, the vector g takes equal values: $g(P') = g(P'') = g(P)$. This implies automatically the condition (γ) for the harmonic vector $H(Q)$ determined by $g(P)$, by the same argument as that given in "GF," § 3.3.

By the condition (γ), symmetric points of $\mathfrak{R}$ give coincident points of M; i.e., the correspondence is univocal between M and the ideal one-sided manifold R which results by identification of symmetric points of $\mathfrak{R}$.

6. The functional $A(g, R)$. As in "GF," our method of solution of the general problem of Plateau is based on the functional

$$A(g, R) = D(H) = \tfrac{1}{2} \int\int (H_u^2 + H_v^2)\, du dv. \tag{5}$$

Here D denotes the "vector Dirichlet integral," defined as half the sum of the Dirichlet integrals of the components of any vector to which it may be applied. Since the harmonic vector H is completely determined by the riemannian manifold R on which it is defined, and by its values g on the boundary C of this manifold, $A(g, R)$ is in effect a functional of the arguments g, R.

For this functional we established in "GF" the following explicit formula,[19] giving it directly in terms of its arguments:

$$(6) \qquad A(g,R) = \frac{1}{8\pi}\int_C\int_C [g(P_1)-g(P_2)]^2 \frac{\partial^2 G(P_1,P_2)}{\partial n_1 \partial n_2}\, ds_1 ds_2.$$

$\partial/\partial n_1, \partial/\partial n_2$ denote differentiation in the direction of the interior normal to C at the points P_1, P_2 respectively.

G denotes the Green's function $G(P,Q)$ of R, which is defined uniquely by the same properties that are specified in § 3 of "GF" for the case where R has a finite number of boundaries and finite connectivity.

As pointed out in connection with formula (3.16) of "GF," the "kernel" of the integral (6) for $A(g,R)$,

$$(7) \qquad G''(P_1,P_2) = \frac{\partial^2 G(P_1,P_2)}{\partial n_1 \partial n_2},$$

is always *positive*.

The integral (6) is improper, due to the singular part $2/\overline{P_1P_2}^2$ of $G''(P_1,P_2)$ when P_2 approaches P_1. Accordingly, we must integrate first with $\overline{P_1P_2} \geqq \epsilon > 0$, obtaining $A_\epsilon(g,R)$, and then define

$$(8) \qquad A(g,R) = \lim_{\epsilon=0} A_\epsilon(g,R).$$

Since the integrand of (6) is positive, the indicated limit always exists definitely either as a finite positive value or as $+\infty$; also, the approach of A_ϵ to A is always in the increasing sense.

We may repeat here the remark of "GF,"[20] that logically the formula (6) can be taken as *definition* of the functional $A(g,R)$, without reference to any equivalence to $D(H)$; therefore our theory is independent of the minimum property of the Dirichlet integral, namely:

$$(9) \qquad D(H) \leqq D(\mathfrak{x}),$$

where $\mathfrak{x}(u,v)$ is any piece-wise continuously differentiable vector function on R with the same boundary values as H.

It is entirely unnecessary to repeat the proof of the transformation of (5) into (6) given for the "algebraic" case in "GF," which depends largely on Green's formula, expressing a double integral over any region as a simple integral over the boundary of that region. For the proof can be carried over almost literally to the general "analytic" case. We only remark that the locus C of each integration variable P_1, P_2 now consists of the totality of real branches of the analytic curve $\mathfrak{A}$ which corresponds to R, and that there may

[19] *Loc. cit.*, formula (3.14).

[20] See the paragraph containing formula (3.18).

be an infinite number of these real branches (see the examples of § 4, except (iii), (vi)).

The range over which $A(g, R)$ is now defined is very wide. Its second argument may be any riemannian manifold whatever in the sense of the general definition given in the formulation of Problem P (§ 2). Its first argument is any vector-function defined on the boundaries of this manifold.

7. The variational theorem.

By applying to the explicit formula (1) for $H(Q)$ the operations next indicated, we easily derive, as in "GF,"[21] the following formula:

$$(10)\qquad \frac{\partial H(Q)}{\partial \xi}\cdot\frac{\partial H(Q)}{\partial \eta} = -\frac{1}{16\pi^2}\int_C\int_C [g(P_1)-g(P_2)]^2 \times \frac{\partial^2}{\partial n_1 \partial n_2}\left\{\frac{\partial G(P_1,Q)}{\partial \xi}\,\frac{\partial G(P_2,Q)}{\partial \eta}+\frac{\partial G(P_1,Q)}{\partial \eta}\,\frac{\partial G(P_2,Q)}{\partial \xi}\right\} ds_1 ds_2.$$

But the most important feature of "GF" is the proof[22] that it is possible to vary the form of the *algebraic* curve $\mathcal{A}$ and the position upon its Riemann surface $\mathcal{S}$ of the points P_1, P_2, i. e., $\mathcal{A}, P_1, P_2$ into $\mathcal{A}_\epsilon, P_1(\epsilon), P_2(\epsilon)$, so that—if Q denote an arbitrarily assigned point of R, and ξ, η any two given perpendicular directions on R at Q—we have for the variation of Green's function of the semi-surface R of $\mathcal{S}$ the formula

$$(11)\qquad \delta G(P_1,P_2) = \frac{\partial G(P_1,Q)}{\partial \xi}\,\frac{\partial G(P_2,Q)}{\partial \eta}+\frac{\partial G(P_1,Q)}{\partial \eta}\,\frac{\partial G(P_2,Q)}{\partial \xi},$$

where δ denotes the variational operator $\partial/\partial\epsilon|_{\epsilon=0}$. This proof was accomplished by giving in a perfectly explicit form the line equation of $\mathcal{A}_\epsilon$ in terms of homogeneous line coördinates u, v, w, where the coefficients were certain simple (linear or constant) functions of the parameter ϵ.

The proof in question goes over practically word for word to the present general case, where $\mathcal{A}$ is any real *analytic* curve, and we may therefore confine ourselves to recasting, with appropriate verbal modifications, the statement of the result, which is as follows.

GENERAL VARIATIONAL THEOREM CONCERNING GREEN'S FUNCTION

THEOREM. *Let $\mathcal{A}$ denote any real analytic curve, on whose Riemann surface $\mathcal{S}$ the points $Q, \bar{Q}$ are any two conjugate imaginary points.*

[21] Formula (3.7).

[22] Compare "GF", § 5.4.

Let the tangent lines $t, \bar{t}$ to $\mathcal{A}$ at $Q, \bar{Q}$ intersect in the real point O. If O is taken as origin of a rectangular axis system, and the position of an arbitrary tangent line to $\mathcal{A}$ is determined by the coefficients u, w in its slope-intercept equation: $y = ux + w$, then the line equation of $\mathcal{A}$ will evidently have the form

$$\mathcal{A}: \qquad (u - u_0)(u - \bar{u}_0)K(u) + wL(u, w) = 0,$$

where $u_0, \bar{u}_0$ are the conjugate imaginary slopes of $t, \bar{t}$, and K, L denote analytic functions or convergent power-series in $u - u_0, w$.

Construct now the family of real analytic curves with real parameter ϵ,

$$\mathcal{A}_\epsilon: \quad [u - \{u_0 + \epsilon(a + bi)\}][u - \{\bar{u}_0 + \epsilon(a - bi)\}]K(u) + wL(u, w) = 0,$$

where $a + bi$, $a - bi$ are any conjugate complex constants.

Then by rectilinear projection from O the Riemann surfaces $\mathcal{S}, \mathcal{S}_\epsilon$ of $\mathcal{A}, \mathcal{A}_\epsilon$ are set into one-one, continuous, and conformal correspondence: P to $P(\epsilon)$, except in the immediate vicinity of $Q, \bar{Q}$. In particular, the real branches C, C_ϵ of $\mathcal{A}, \mathcal{A}_\epsilon$ are thereby set into one-one continuous correspondence. This depends on the circumstance that all the real tangents from O to $\mathcal{A}_\epsilon$, whose slopes are the real roots of the equation

$$K(u) = 0,$$

remain invariant, since this equation is independent of ϵ.

In fact, for the same reason, all the tangents, real and imaginary, from O to $\mathcal{A}_\epsilon$ except $t_\epsilon, \bar{t}_\epsilon$ remain fixed. These, however, vary with ϵ; and t_ϵ, for instance, intersects the Riemann surface $\mathcal{S}$ of $\mathcal{A}$ in two points[23] *$Q'_\epsilon, Q''_\epsilon$, near to Q, which, as ϵ passes through the value zero from positive to negative, always enter Q from two opposite directions α', α'' and leave in the perpendicular opposite directions β', β''. Let the angle bisectors of $\alpha'\beta', \alpha'\beta''$ be the perpendicular directions ξ, η.*

Denote by $G(P_1, P_2)$ the Green's function of either conjugate semi-surface R of the Riemann surface $\mathcal{S}$.[24] *This is for the case where $\mathcal{S}$ is orthosymmetric; if $\mathcal{S}$ is diasymmetric, $G(P_1, P_2)$ will denote the Green's function of $\mathcal{S}'$, which is $\mathcal{S}$ severed along its transition curves C, the real branches of $\mathcal{A}$.*

Then under the precedingly described variation of the form of $\mathcal{A}$ and of the position upon its Riemann surface $\mathcal{S}$ of the points P_1, P_2, we have for Green's function the variational formula

---

[23] We suppose that $t, \bar{t}$ have ordinary two-point contact with $\mathcal{A}$ at $Q, \bar{Q}$—higher contact can always be avoided by a preliminary birational transformation.

[24] The notation is supposed arranged so that R contains Q, while its conjugate semi-surface R' contains $\bar{Q}$.

$$(11)\quad \delta G(P_1, P_2) = \frac{\partial G(P_1, Q)}{\partial \xi}\,\frac{\partial G(P_2, Q)}{\partial \eta} + \frac{\partial G(P_1, Q)}{\partial \eta}\,\frac{\partial G(P_2, Q)}{\partial \xi}$$

(*apart from an inessential numerical factor*).

Finally, the directions ξ, η can be made to coincide with any preassigned perpendicular directions at Q by proper choice of the constant $a : b$.

Figure 5 of "GF" may be used again here for illustration.[25]

8. Fundamental minimum principle. By combining the three fundamental formulas of this paper, namely, (6), (10), (11), we obtain the following:

$$(12)\quad \delta A(g, R) = -2\pi \frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta}.$$

This depends on the validity of performing the operation δ under the integral sign, and on the commutativity of δ with the differential operator $ds_1 ds_2\, \partial^2/\partial n_1 \partial n_2$, both of which are readily established.[26]

Let R denote any riemannian manifold of the topological type $\mathfrak{T}$ prescribed for the required minimal surface M in the Plateau problem, and let $\mathfrak{x} = g(P)$ be the parametric, or topological representation of the boundary Γ assigned for M on the boundary C of R. Then we shall term the composite entity (g, R) a *representation* of Γ. For each such representation the positive-valued functional $A(g, R)$ can be formed.

We shall use the notation $d(\Gamma, \mathfrak{T})$ for the *lower bound* of $A(g, R)$ when (g, R) varies over all proper representations of the given boundary Γ:

$$(13)\quad d(\Gamma, \mathfrak{T}) = \min A(g, R),$$

where the notation "min" does not prejudice the question of attainment of the lower bound. For any particular data $(\Gamma, \mathfrak{T})$, d may be a finite positive quantity or $+\infty$.

Let us suppose that $d(\Gamma, \mathfrak{T})$ is finite, and that its value as lower bound is *attained* for a certain particular representation:

$$(14)\quad A(g^*, R^*) = d(\Gamma, \mathfrak{T}) = \min.$$

Now perform on the Riemann surface R^*, or its equivalent real analytic curve $\mathcal{A}$, the construction described in the Variational Theorem, which con-

---

[25] In the proof of the preceding Variational Theorem, an essential step is to cause the arbitrarily chosen point Q of R to become a branch-point of a conformally equivalent Riemann surface R'. If R is represented as covering surface over the z-plane, and Q lies over z_0, this requirement can be accomplished by means of the transformation $z' = (z - z_0)^2$. Then the branch-point Q can be displaced as in "GF," § 5. 1.

[26] See the discussion in "GF" associated with formula (4. 10).

verts R^* point-wise into R_ϵ. The correspondence between the real branches C^*, C_ϵ of or boundaries of R^*, R_ϵ, is one-one and continuous. Therefore if we transfer the value $\mathfrak{x}$ of $g(P)$ from each point P of C^* to the corresponding point $P(\epsilon)$ of C_ϵ, we again have a *representation* (g_ϵ, R_ϵ) of the assigned boundary Γ. It follows that the function of ϵ,

$$J(\epsilon) = A(g_\epsilon, R_\epsilon) \tag{15}$$

has a minimum value for $\epsilon = 0$, and therefore

$$J'(0) = \delta A(g, R) = 0. \tag{16}$$

Consequently, by (12),

$$\frac{\partial H(Q)}{\partial \xi} \cdot \frac{\partial H(Q)}{\partial \eta} = 0; \tag{17}$$

but this is precisely the characteristic condition for a minimal surface.

In sum, we have the following

Minimum principle. *Among all representations (g, R) of a given boundary Γ where R is of a prescribed topological type $\mathfrak{T}$, let the minimum value of $A(g, R)$ be attained for a particular representation (g^*, R^*). Then the harmonic vector on R^*,*

$$\mathfrak{x} = H(Q),$$

determined by the values $g^*(P)$ on the boundary C^*, defines a minimal surface M of topological type $\mathfrak{T}$ bounded by Γ; in other words, $\mathfrak{x} = H(Q)$ solves the Plateau problem for the given boundary Γ and prescribed topological type $\mathfrak{T}$.

9. Attainment of the minimum of $A(g, R)$. All that now remains is to establish the *attainment* of the minimum of $A(g, R)$.

The proof is the same for the general "analytic" case, here under consideration, as for the "algebraic" case of "GF" (§ 8). We have only to verify that the Fréchet postulates are obeyed by our fundamental minimum problem:

$$A(g, R) = \min. \tag{18}$$

First, the set of all riemannian manifolds R of a given topological type $\mathfrak{T}$ is *compact*. For these can be realized by means of normal polygons in the N. E. plane, and the set of all normal polygons, since it depends on a denumerable infinity of parameters, is compact; the same is therefore true of its subset of type $\mathfrak{T}$. In fact, we can apply the Bolzano-Weierstrass condensation principle to each parameter separately; its validity for the denumerable infinity then follows by the standard "diagonal process."

As to g, it is defined by a finite or infinite number of monotonic functions, which give the one-one continuous correspondence between each real branch of the analytic curve $\mathcal{A}$ related to R and the corresponding component contour of the assigned total boundary Γ. It is a standard theorem that any set of uniformly bounded monotonic functions is compact. Accordingly, it results again by the diagonal process that the set of all parametric representations g of a given boundary Γ in n-space on the boundary C of a riemannian manifold R is compact.

Second, the functional $A(g, R)$ is lower semi-continuous, due to its representation, according to (8), as the *increasing* limit of the *continuous* functional $A_\epsilon(g, R)$.

What requires special provision is the *closure* of the set of all representations (g, R) of Γ. For a limit of proper representations may not be such a representation. Indeed, the limit of a variable riemannian manifold R of a specified topological type $\mathcal{T}$ may not belong to this type, but may separate into a finite or infinite number of disconnected parts, or, without separating, may reduce to a lower topological structure through the disappearance of a number of handles or branch-cuts; or both these phenomena may be combined. Again, R may be improper as to its boundary C, which while R varies may always be, as is required, topologically equivalent to Γ, but in the limit may lose this property; for instance, two disjoint components of C may approach to one another and finally touch in the limit,[27] whereas the corresponding component contours of Γ remain fixed and disjoint.

As to g, the continuous monotonic functions by which it is defined may approach in the limit to improper monotonic functions, with intervals of constancy or with discontinuities in the form of finite jumps. The geometric interpretation is that a *partial* arc[28] of a real branch C of the analytic curve $\mathcal{A}$ may correspond to a single point on one of the contours Γ (continuously improper), or *vice versa* (discontinuously improper); an extreme case is where an entire real branch C of $\mathcal{A}$ corresponds to a single point of the related contour Γ, while this entire contour corresponds to a single point of the real branch (degenerate representation).

We now have a complete survey of the improper representations (g, R), which may be listed as follows.

(I) R improper:

(Ia) R separates or reduces to a lower topological type than the prescribed one $\mathcal{T}$.

---

[27] This gives a real double point of $\mathcal{A}$, as pointed out in "GF", § 8.2 (see Ib).

[28] That this arc is less than the entire real branch is what distinguishes this type of improper representation from the completely degenerate type.

(Ib) The boundary C of R is no longer homeomorphic with the boundary Γ assigned for the required minimal surface, in that two component closed curves of C touch whereas the corresponding contours of Γ do not; i. e., $\mathcal{A}$ acquires a real double point.

(II) g improper:

(IIa) Discontinuously improper.

(IIb) Continuously improper.

(IIc) Degenerate.

The closure of the set of all representations (g, R) of Γ is secured only at the expense of including all these improper forms of representation.

This being done, all the Fréchet postulates are obeyed by the functional $A(g, R)$, which therefore *attains its minimum value, but perhaps for an improper representation.*

The point is now to exclude the possibility that the minimizing representation (g^*, R^*) be improper.

10. Sufficient conditions for the solution. We may think of the reduced form (Ia) of R as follows. Let R_0 denote any particular riemannian manifold of the type $\mathfrak{T}$, and $\mathcal{A}_0$ the corresponding analytic curve. Imagine the real analytic curve $\mathcal{A}$ to vary continuously starting with the form $\mathcal{A}_0$. Then so long as $\mathcal{A}$ does not become reducible—i. e., the equation of $\mathcal{A}$: $F(x, y) = 0$, does not separate into factors—and so long as $\mathcal{A}$ does not acquire any new double points,[29]—either imaginary ones in conjugate pairs, or real ones,—$\mathcal{A}$ and its equivalent R remain in the class $\mathfrak{T}$. However, in the continuous variation of $\mathcal{A}$, as soon as we encounter reducibility or new double points, $\mathcal{A}$ or R belongs to the improper type.

We shall denote by $\mathfrak{T}'$ the class of all analytic curves or equivalent riemannian manifolds R', which are thus obtainable as limits of manifolds of type $\mathfrak{T}$ without being of this type themselves. $\mathfrak{T}'$ may be said to constitute the "frontier" of $\mathfrak{T}$ in the "space" of all riemannian manifolds.

By formula (5), the definition of $A(g, R)$ is easily extended to improper forms of representation (g', R'). In case R' separates, we simply form the sum of the appropriate Dirichlet integrals for the individual component manifolds.

Let us denote by $d(\Gamma, \mathfrak{T}')$ the lower bound of $A(g', R')$, formed only for the improper representations:

$$d(\Gamma, \mathfrak{T}') = \min A(g', R'). \tag{19}$$

[29] Or multiple points. "New" means not resulting by following out the continuous variation of the multiple points of $\mathcal{A}_0$.

A simple form of reasoning establishes that between this and the similar lower bound for the proper type $\mathfrak{J}$ the following relation of inequality holds in every case:

$$d(\Gamma, \mathfrak{J}) \leqq d(\Gamma, \mathfrak{J}'). \tag{20}$$

In fact, let (g', R') denote any improper representation, and (g_m, R_m) an approaching sequence of improper representations. Then, while we have in any event the relation of lower semi-continuity:

$$A(g', R') \leqq \liminf_{m\to\infty} A(g_m, R_m), \tag{21}$$

it is, indeed, always possible to arrange the approaching sequence in a special way so that we have the relation of continuity:

$$A(g', R') = \lim_{m\to\infty} A(g_m, R_m). \tag{22}$$

For instance, if R' separates, so that the corresponding analytic curve $\mathcal{A}'$ is reducible:

$$F_1(x, y)F_2(x, y) = 0, \tag{23}$$

we can form the family of irreducible analytic curves $\mathcal{A}_\epsilon$:

$$F_1(x, y)F_2(x_1, y) = \epsilon, \tag{24}$$

and also set the real branches of $\mathcal{A}$ and $\mathcal{A}_\epsilon$ into one-one continuous correspondence, P to $P(\epsilon)$, in a way that depends analytically on the position of the point P and the parameter ϵ. Then let us transfer the value of $g'(P)$ from each point P to the corresponding point $P(\epsilon)$. We obtain in this way a family of representations (g_ϵ, R_ϵ), proper for sufficiently small $\epsilon > 0$, and including the given improper representation (g', R') for $\epsilon = 0$. By formula (6), we can see that $A(g_\epsilon, R_\epsilon)$ is a continuous function of ϵ, including the end-value $\epsilon = 0$; for Green's function $G(P_1, P_2)$ depends analytically on ϵ. Giving then to ϵ a sequence of values ϵ_m tending to zero, we have the desired relation (22).

This implies that, given any improper representation (g', R'), we can always find a proper one (g, R) such that $A(g, R)$ differs arbitrarily little in value from $A(g', R')$. The stated relation (20) is an immediate consequence.

Suppose now that for particular data $(\Gamma, \mathfrak{J})$ *we have, besides the natural assumption,*

$$d(\Gamma, \mathfrak{J}) < +\infty, \textit{ i. e., finite}, \tag{25}$$

also the strict form of inequality in (20):

$$d(\Gamma, \mathfrak{J}) < d(\Gamma, \mathfrak{J}'). \tag{26}$$

Then it follows that every improper representation (g', R') of type (Ia) is excluded as a possibility for the minimizing representation of $A(g, R)$, since

$$A(g', R') \geqq d(\Gamma, \mathfrak{J}') > d(\Gamma, \mathfrak{J}). \tag{27}$$

All the other forms of improper representation can also be excluded, just as in "GF," namely: (Ib) because then $A(g, R) = +\infty$, contrary to (25); (IIa) for the same reason; (IIb) because inconsistent with the variational condition (23) expressing a conformal correspondence between R and the required minimal surface M; (IIc) because the degenerate representations behave analytically a good deal like the case (Ia) of separation of R—indeed, degeneration of the representation of a particular contour Γ_1 of the given system Γ amounts in the formulas to the detachment from M of the simply-connected minimal surface bounded by Γ_1, so that, as a consequence, we have for any degenerate representation also the condition (27).

In sum, the representation (g^*, R^*) for which the minimum value of $A(g, R)$ is attained:

$$A(g^*, R^*) = d(\Gamma, \mathfrak{J}) = \min., \tag{28}$$

is a proper one.

With this, we have completed the solution of the most general form of the problem of Plateau stated at the beginning as Problem P, and may now formulate our result in the following theorem.

THEOREM I. *For particular data* $(\Gamma, \mathfrak{J})$, *let* $d(\Gamma, \mathfrak{J})$ *be finite, while between* $d(\Gamma, \mathfrak{J})$ *and* $d(\Gamma, \mathfrak{J}')$ *the strict form of inequality prevails:*

$$d(\Gamma, \mathfrak{J}) < d(\Gamma, \mathfrak{J}').$$

Then there exists a minimal surface M *of topological type* $\mathfrak{J}$ *bounded by* Γ.

11. Least area. Effective as we have just seen the functional $A(g, R)$ to be in the analytic treatment of the Plateau problem, there is no doubt that, from the standpoint of concreteness and interest, the most important feature of the problem resides in the least area property of the surface whose existence it requires us to establish.

This least area property of M we have proved many times before in previous papers on the basis of the conformal mapping of polyhedral surfaces on some standard riemannian manifold. The polyhedral surfaces are those which figure in the Lebesgue definition of area:

$$\mathfrak{A}(S) = \min \lim_{m \to \infty} \inf \mathfrak{A}(\Pi_m), \tag{29}$$

where the minimum is with respect to all sequences of polyhedra Π_m which tend to the surface S in the sense of Fréchet.

In the present general case, where the topological type of S is perfectly arbitrary, we may suppose every approaching polyhedron Π_m to be of this type, or else we may suppose Π_m to be of a variable topological type $\mathfrak{T}_m$ which—in a sense not hard to define by reference to the corresponding normal polygons $\mathfrak{P}_m$ on which the polyhedra are mapped conformally—tends to the type $\mathfrak{T}$ of S. The possibility of this conformal map of Π_m on $\mathfrak{P}_m$ follows from the Koebe uniformization theory just as well in the present case, where Π_m may have an infinite number of faces and boundaries and an infinite type of connectivity, as in the cases discussed in previous papers, where all these things were finite.

We define the lower bound of areas:

$$a(\Gamma, \mathfrak{T}) = \min \mathfrak{A}(S) \tag{30}$$

where S ranges over all surfaces of type $\mathfrak{T}$ bounded by Γ.

The fundamental result, quite readily provable by the methods just referred to, is *the equality of this functional a based on area to the functional d based on* $A(g, R)$:

$$a(\Gamma, \mathfrak{T}) = d(\Gamma, \mathfrak{T}). \tag{31}$$

Furthermore, we have for the area of M:

$$\mathfrak{A}(M) = d(\Gamma, \mathfrak{T}). \tag{32}$$

For let us consider the integral formula for the area of M:

$$\mathfrak{A}(M) = \int\int \sqrt{EG - F^2}\, du dv, \tag{33}$$

in comparison with the minimum value of $A(g, R)$:

$$A(g^*, R^*) = \int\int \tfrac{1}{2}(E + G)\, du dv = d(\Gamma, \mathfrak{T}). \tag{34}$$

Here E, F, G denote the fundamental quantities of the surface M:

$$E = H_u^2, \qquad F = H_u \cdot H_v, \qquad G = H_v^2, \tag{35}$$

whereby the equivalence of (34) to the defining formula (5) is immediate.

According to the variational condition (17) obeyed by the minimizing representation (g^*, R^*), the correspondence defined by $H(u, v)$ between R and M is conformal; this means that

$$E = G, \qquad F = 0. \tag{36}$$

But under this condition the expressions (33) and (34) are obviously equal; thus we obtain (32).

Combining this with (31), we have the desired least area property of M.

THEOREM I'. *The minimal surface M whose existence is established by Theorem I has the least possible area for its boundaries Γ and topological type $\mathfrak{T}$.*

Similar to (30), *we may define the lower bound*

$$a(\Gamma, \mathfrak{T}') = \min \mathfrak{A}(S'), \tag{37}$$

where S' ranges over all surfaces of the reduced or separating type $\mathfrak{T}'$ bounded by Γ. Like (30), *we have, with a similar proof,*

$$a(\Gamma, \mathfrak{T}') = d(\Gamma, \mathfrak{T}'). \tag{38}$$

These relations enable us to restate the sufficient conditions of Theorem I in the following form:

$$a(\Gamma, \mathfrak{T}) < + \infty, \tag{39}$$

$$a(\Gamma, \mathfrak{T}) < a(\Gamma, \mathfrak{T}'), \tag{40}$$

more concrete and satisfactory from the geometric standpoint.

12. The case of infinite $d(\Gamma, \mathfrak{T})$. In this concluding section we lift the first restriction of Theorem I, and extend our theory to the case where

$$d(\Gamma, \mathfrak{T}) = + \infty. \tag{41}$$

According to (31), this means that Γ is incapable of bounding any surface of type $\mathfrak{T}$ having a finite area, but rather every surface S belonging to the data $(\Gamma, \mathfrak{T})$ has infinite area.

To secure the existence of the required minimal surface M, we need a condition to replace the second one of Theorem I, namely, (26), both of whose terms are now infinite.

Let Γ_m denote any given boundary and $\mathfrak{T}_m$ any topological type of surface that may be bounded by Γ_m. Then in the case where $d(\Gamma_m, \mathfrak{T}_m)$ is finite, we define

$$e(\Gamma_m, \mathfrak{T}_m) = d(\Gamma_m, \mathfrak{T}'_m) - d(\Gamma_m, \mathfrak{T}_m) \geqq 0, \tag{42}$$

i. e., an *essentially non-negative* quantity by (20).

Given now any $(\Gamma, \mathfrak{T})$, we can always express these data as the limit for $m \to \infty$ of a sequence $(\Gamma_m, \mathfrak{T}_m)$ such that $d(\Gamma_m, \mathfrak{T}_m)$ is finite for each value of m. For instance, we can take the topological type $\mathfrak{T}_m$ as characterized for each value of m by a finite number k_m of boundaries and a finite connectivity c_m—where, if necessary, k_m and c_m tend to ∞ with m; while the k_m boundaries Γ_m may be taken as polygons.

This enables us to define, whether $d(\Gamma, \mathfrak{T})$ is finite or infinite,

$$\bar{e}(\Gamma, \mathfrak{T}) = \max \limsup_{m\to\infty} e(\Gamma_m, \mathfrak{T}_m), \tag{43}$$

the maximum being with respect to all possible sequences $(\Gamma_m, \mathfrak{T}_m)$ approaching to $(\Gamma, \mathfrak{T})$ so that $d(\Gamma_m, \mathfrak{T}_m)$ is always finite. According to the remark following (42), *$\bar{e}$ is also an essentially non-negative quantity:*

$$\bar{e}(\Gamma, \mathfrak{T}) \geqq 0 \tag{44}$$

in every case.

Suppose in a particular instance that $\bar{e}$ is strictly positive:

$$\bar{e}(\Gamma, \mathfrak{T}) > 0. \tag{45}$$

Then we state the following theorem, whose proof resembles that already given in the author's previous papers for the case of a finite number of boundaries and finite connectivity, with sufficient closeness that its repetition here is unnecessary.

THEOREM II. *If $d(\Gamma, \mathfrak{T})$ is infinite, and $\bar{e}(\Gamma, \mathfrak{T})$, always non-negative, is strictly positive for the particular data $(\Gamma, \mathfrak{T})$, then a minimal surface M exists of topological type $\mathfrak{T}$ bounded by Γ.*

The condition (45) operates to prevent the degeneration of M, playing in this respect the rôle of the condition (26) in the case of finite $d(\Gamma, \mathfrak{T})$.

The area of M is infinite, as well as that of every surface of type $\mathfrak{T}$ bounded by Γ; hence there can be no question of a least area property for M as a whole. However, a least area property does subsist, so to say, throughout the *interior* of M. We have, in fact, the following theorem, whose proof is again sufficiently similar to that for the "finite" case as to require no repetition.

THEOREM II'. *The area of the minimal surface M whose existence is established in Theorem II is infinite, as well as that of every surface of topological type $\mathfrak{T}$ bounded by Γ. But every region M_1 of M which, with its boundary Γ_1, is completely interior to M has an area which is finite and a minimum for its own boundary Γ_1 and topological type $\mathfrak{T}_1$.*

Therewith, the most general form of the problem of Plateau is solved also for the case of infinite $d(\Gamma, \mathfrak{T})$ or $a(\Gamma, \mathfrak{T})$.

BROOKLYN, N. Y.

HARMONIC FUNCTIONS IN DOMAINS WITH MULTIPLE BOUNDARY POINTS.*

By JOHN W. GREEN.

In a recent paper Maria and Martin [1] discussed the possibility of representing any positive harmonic function $u(P)$ in a domain T by a Stieltjes integral of the form

$$u(P) = \int_t f(S, P)\, d\mu(e_s)\,; \tag{1}$$

where $\mu(e)$ represents a distribution of positive mass on the boundary t of the domain, and $f(S, P)$ is defined for P in T and S on t, and is the same for all $u(P)$. In order that this representation be possible, it was found necessary to impose certain conditions on the domain. In a treatment of the Dirichlet problem for domains with boundary points counted multiply, Perkins [2] has employed the idea of a "boundary element" as a fundamental concept instead of boundary point. The papers of Perkins and of Maria and Martin are both generalizations of certain results of G. C. Evans [3] which apply only to the case of two dimensions. The present paper is concerned with the adaptation of Perkin's concept, by means of the introduction of an abstract space, to the representation of harmonic functions in a domain by means of a Stieltjes integral with respect to a function of sets of boundary elements. The representation will apply to the extension of both the Dirichlet problem and the results of Maria and Martin to a wider class of domains than has heretofore been possible. For any facts concerning boundary elements stated but not proved in this paper, the reader is referred to the paper of Perkins quoted.

1. Introduction. Let T be a bounded domain situated in any Euclidean space, and let t be the set of all its boundary points. The following theory is valid for any number of dimensions, although the language will be frequently for convenience that of three dimensions. We consider with Perkins boundary

* Received June 5, 1938; revised September 29, 1938.

[1] Alfred J. Maria and Robert S. Martin, "Representation of positive harmonic functions," *Duke Mathematical Journal*, vol. 2, no. 3, pp. 517-529.

[2] F. W. Perkins, "The Dirichlet problem for domains with multiple boundary points," *Transactions of the American Mathematical Society*, vol. 38, no. 1, pp. 106-144.

[3] G. C. Evans, *The Logarithmic Potential*, New York, 1927, Chaps. V and VI.

elements of T. A boundary element γ of T is defined to be a sequence of partial domains of T: $T_1, T_2, \cdots$ with boundaries $t_1, t_2, \cdots$ with the properties:

(a) $T_n \supset T_{n+1}$.

(b) Each t_n contains at least one point of t.

(c) $t_{n+1} - t \cdot t_{n+1}$, is at a positive distance from $t_n - t \cdot t_n$.

(d) The diameter of T_n approaches zero with $1/n$.

The set of points $t_n - t \cdot t_n$ together with its limit points is called the auxiliary boundary of T_n. (c) implies that the auxiliary boundaries of T_n and T_{n+1} have no points in common. Two boundary elements γ: $T_1, T_2, \cdots$ and $\bar{\gamma}$: $\bar{T}_1, \bar{T}_2, \cdots$ are considered identical if given n, one can find m such that

$$\begin{aligned} T_n &\supset \bar{T}_{m+k}, \\ \bar{T}_n &\supset T_{m+k}, \end{aligned} \qquad (k = 0, 1, 2, \cdots).$$

One of these two conditions can be shown to imply the other.

Every boundary element contains a point P of t; that is, there is one and only one point P of t on all t_n. On the other hand, given a point P of t, there may possibly be no boundary element containing it. For example, let T be the plane domain represented in rectangular coördinates by points (x, y) such that $0 < x < 1$, $0 < y < 1$, except that slits of length $\frac{1}{2}$ parallel to the y axis and starting at the points $(1/2^k, 0)$, $k = 1, 2, \cdots$, are deleted. The strip $\mathfrak{L}$ of points with coördinates $x = 0$, $0 \leq y < \frac{1}{2}$, belongs to t, but these points are not contained in boundary elements of T. In fact, if P is one of these points, there is no domain contained in T and of arbitrary small diameter having P for boundary point.

As an example of an opposite extreme, consider the domain consisting of the interior of the unit circle about the origin O of the x, y plane. From this domain delete slits of length $1/2^k$ starting at O and making angles with Oy of $\pm (\pi/2) \cdot (m/2^k)$, $m = 0, 1, \cdots, 2^k$, $k = 1, 2, \cdots$. The resulting domain has non-denumerably many boundary elements containing O.

Let $\gamma = \{T_i\}$ be an element of the boundary of T and let $S(P, \rho)$ be the sphere of radius ρ and with center at the point P of t contained in γ. Let T_k be the first domain of the sequence $\{T_i\}$ to be contained in $S(P, \rho)$; then $\mathfrak{S}(\gamma, \rho)$, the pseudo-spherical domain at γ of radius ρ, is defined to be the set of all points of T in $S(P, \rho)$ which can be joined to points of T_k by continuous curves lying in T and in $S(P, \rho)$. $\mathfrak{S}(\gamma, \rho)$ is independent of the particular sequence $\{T_i\}$ defining γ. Any sequence $\mathfrak{S}(\gamma, \rho_i)$ with ρ_i approaching zero monotonically may be used as the sequence of domains defining γ. If γ is a

point of T, then $\mathfrak{S}(P, \rho)$ is defined to be the set of points of T which can be joined to P by continuous curves lying in T and within the sphere of radius ρ at P. A pseudo-spherical domain is said to contain an element γ if it contains all pseudo-spherical domains of γ of sufficiently small radius. A sequence of points or boundary elements of T is said to approach γ if all but a finite number of the points or boundary elements respectively are contained in any $\mathfrak{S}(\gamma, \rho)$.

2. Definition and properties of a certain abstract space. Consider the points P of T and the boundary elements γ of T as forming a space $\mathfrak{J}$. Let π denote an element of this space, either a γ or a P, and let π' denote another such element.[4] The distance (π, π) is defined as follows: Let ρ be the lower bound of $\bar{\rho}$ such that $\mathfrak{S}(\pi, \bar{\rho})$ contains π' and ρ' the lower bound of $\bar{\rho}$ such that $\mathfrak{S}(\pi', \bar{\rho}]$ contains π. Then

$$(\pi, \pi') = \tfrac{1}{2}(\rho + \rho').$$

Obviously $(\pi, \pi') = (\pi', \pi)$. If $\mathfrak{S}(\pi, \rho)$ contains π', then $\mathfrak{S}(\pi', 2\rho)$ contains all of $\mathfrak{S}(\pi, \rho)$ and so contains π. Consequently we have

$$\rho_1 \leq 2\rho \quad \text{and similarly} \quad \rho \leq 2\rho'.$$

From this it follows that if $\mathfrak{S}(\pi, \rho)$ contains π', then $(\pi, \pi') < 2\rho$. If π is not identical with π', then (π, π') is different from zero. For, if the contrary were true, then π and π' would obviously have to contain the same point P. But if π and π' both contain P, it is clear that $\mathfrak{S}(\pi, \rho)$ and $\mathfrak{S}(\pi', \rho)$ are either identical or have no point in common. If they were identical for all ρ, then π and π' would be the same, for any element can be considered as the set of its pseudo-spherical domains. Thus for ρ_1 sufficiently small, $\mathfrak{S}(\pi, \rho_1)$ and $\mathfrak{S}(\pi', \rho_1)$ have no point in common, and (π, π') is not less than ρ_1.

Let now π, π', and π'' be three distinct elements of $\mathfrak{J}$. Let $\mathfrak{S}(\pi, \bar{\rho}_1)$ contain π' and $\mathfrak{S}(\pi', \bar{\rho}_2)$ contain π''. Then $\mathfrak{S}(\pi, \bar{\rho}_1 + \bar{\rho}_2)$ contains all of $\mathfrak{S}(\pi', \bar{\rho}_2)$ and so contains π''. Consequently, if ρ_1 and ρ_2 are the lower bounds of $\bar{\rho}_1$ and $\bar{\rho}_2$ respectively, and ρ_3 the lower bound of $\bar{\rho}_3$ such that $\mathfrak{S}(\pi, \bar{\rho}_3)$ contains π'', we have $\rho_1 + \rho_2 \geq \rho_3$. Similarly $\rho'_1 + \rho'_2 \geq \rho'_3$, where ρ'_3 is the lower bound of $\bar{\rho}_3$ such that $\mathfrak{S}(\pi'', \bar{\rho}_3)$ contains π, etc. From this follows immediately the triangular inequality,

$$(\pi, \pi'') \leq (\pi, \pi') + (\pi', \pi'').$$

---

[4] When there is no cause for confusion, a single letter P or γ will be used to represent a point of T or boundary element of t and the element of the space $\mathfrak{J}$ corresponding.

2.1. THEOREM. *The space $\mathfrak{J}$ is complete.*

Let $\{\pi_n\}$ be a convergent sequence of elements of $\mathfrak{J}$. These correspond either to points of T or to elements of the boundary of T. If infinitely many of them correspond to points $\{P_n\}$ of T which remain a positive distance $\geq \delta > 0$ from t, it is obvious that there is a point P of T to which $\{P_n\}$ converges in the ordinary sense, and that this point corresponds to the limit element π of $\{\pi_n\}$. The cases that offer a problem then, are those in which there is a sequence involving infinitely many boundary elements, or infinitely many points of T converging to a point of t. The former case can be reduced to the latter by considering an auxiliary sequence $\{P_n\}$ of points of T determined sufficiently close to the elements $\{\pi_n\}$. The sequence $\{P_n\}$ will be convergent and if it possesses a limit element, $\{\pi_n\}$ will possess the same limit element because of the triangular inequality.

Let then $\{P_n\}$ be a sequence of points which, interpreted as elements of $\mathfrak{J}$, form a convergent sequence in terms of the metric of that space. It is obvious that the points $\{P_n\}$ also form a convergent sequence in the ordinary sense. Let P be the limit point of $\{P_n\}$; we are supposing that P is on t. Take a sequence of values $\rho_1 > \rho_2 > \cdots$ tending to zero and consider the sphere $S(P, \rho_1)$. The portion of this sphere in common with T consists of finitely or denumerable infinitely many domains. All but finitely many of the P_n lie in a single one of these domains. For suppose the contrary. Since the P_n tend to P, it will be possible to find n' such that for $n \geq n'$, $\overline{PP_n} < \rho_1/4$. Also since $(P_n, P_{n+k}) \to 0$, it will be possible to find n'' such that for $n \geq n''$, $\mathfrak{S}(P_n, \rho_1/4) \supset P_{n+k}$. We can find, then, P_i in T_i and P_j in T_j, where T_i and T_j are two different domains determined by T and $S(P, \rho_1)$, such that $\overline{PP_i} < \rho_1/4$, and $\mathfrak{S}(P_i, \rho_1/4)$ contains P_j. This means that P_i and P_j can be joined in T by a curve in $S(P_i, \rho_1/4)$; this curve cannot depart from $S(P, \rho_1)$, and so T_i and T_j must be the same domain, which is contrary to assumption. Let then T_1 be the domain of $S(P, \rho_1)$ which contains all P_n for $n \geq n_1$. Similarly, let T_2 be the domain of $S(P, \rho_2)$ which contains all P_n for $n \geq n_2$. T_2 is contained in T_1. In this manner we form the nested sequence of domains: $T_1 \supset T_2 \supset T_3 \cdots$ determining a boundary element γ. For $n \geq n_i$, $\mathfrak{S}(\gamma, \rho_i) \supset P_n$. We see that in terms of the metric of $\mathfrak{J}$, $(\gamma, P_n) \leq 2\rho_i$ for $n \geq n_i$, and $\{P_n\}$ converges to γ.

2.2. We now formulate the following condition on the domain S.

CONDITION (A). *Let P be any point of t, and let $S(P, \rho)$ be the sphere at P of radius ρ. Then there exists $\delta > 0$, $\delta = \delta(P, \rho)$, such that of the*

finitely or denumerably infinitely many domains forming the set common to T and to $S(P, \rho)$, all but a finite number are at distances $\geq \delta$ from P.

This condition excludes the first example described in § 1. Subsequently some domains not satisfying condition (A) will be readmitted to the discussion provided certain other conditions are satisfied, and among these domains will be the example mentioned. The reason for imposing condition (A) will later be made obvious. Unless otherwise stated, the domains henceforth considered will satisfy (A).

2.3. THEOREM. *The space $\mathfrak{J}$ is compact.*

Let $\{\pi_n\}$ be an infinite sequence of elements of $\mathfrak{J}$. In this sequence there are either infinitely many elements corresponding to points of T, or infinitely many corresponding to elements of the boundary of T. Let us suppose the former, the arguments in the other case being exactly analogous.

Let then $\{P_n\}$ be an infinite sequence of points of T. From this sequence we can select a subsequence, which we denote again by $\{P_n\}$, which converges to some point P. Unless P is on t, it is obviously true that eventually $(P_n, P) = \overline{P_n P}$, and so $\lim_{n=\infty} (P_n, P) = 0$. We assume then that P is on t; our aim is to show the existence of a boundary element γ such that $\lim_{n=\infty} (\gamma, \bar{P}_n) = 0$, for a properly chosen subsequence $\{\bar{P}_n\}$. Let $\rho_1 > \rho_2 > \cdots$ be a sequence of values tending to zero, and consider $S(P, \rho_1)$. Let n_1 be such that for $n \geq n_1$, $\overline{PP}_n < \delta(P, \rho_1)$. Then by condition (A), all P_n for $n \geq n_1$ lie in a finite number N_1 of the domains forming the set common to T and to $S(P, \rho)$. One of these domains, say T_1, contains infinitely many points of $\{P_n\}$. Next consider $S(P, \rho_2)$. We obtain similarly that there exists T_2 contained in $S(P, \rho_2)$ and in T_1 such that T_2 contains infinitely many points of $\{P_n\}$. Continuing in this manner, we obtain a nested sequence of domains: $T_1 \supset T_2 \supset T_3 \cdots$ defining a boundary element γ. To construct $\{\bar{P}_n\}$ we select out of the sequence $\{P_n\}$, $\bar{P}_1$ in T_1, P_2 in $T_2, \cdots$. Then $\mathfrak{S}(\gamma, \rho_n) \supset \bar{P}_n$, or $(\gamma, \bar{P}_n) < 2\rho_n$, and the sequence $\{\bar{P}_n\}$ converges to the element γ of $\mathfrak{J}$. Thus $\mathfrak{J}$ is compact.

We have seen that condition (A) implies the compactness of $\mathfrak{J}$. Condition (A) is also the least restrictive condition making this implication, for from the compactness of $\mathfrak{J}$, condition (A) follows. In fact, suppose that $\mathfrak{J}$ is compact and that for P and ρ, condition (A) is not satisfied. Then from the domains forming the set common to T and to $S(P, \rho)$ we can select $T_1, T_2, T_3, \cdots$ such that the distance of T_n from P is less than $\rho/2^n$. Select P_n in T_n such that $\overline{PP}_n < \rho/2^n$. Then this sequence $\{P_n\}$ has no subsequence

converging to a boundary element. For suppose such a subsequence does exist converging to some γ. This γ will obviously have to contain the point P. Consider $\mathfrak{S}(\gamma, \rho/2)$. If $\mathfrak{S}(\gamma, \rho/2)$ contains P_n, it cannot contain P_{n+k}, for such would imply that P_n and P_{n+k} could be joined in T by a curve in $\mathfrak{S}(\gamma, \rho/2)$. This curve cannot depart from $S(P, \rho)$, and P_n and P_{n+k} would lie in the same domain of $S(P, \rho)$. The assumption that condition (A) does not hold for T leads then to a contradiction of the hypothesis that $\mathfrak{F}$ is compact.

For the analysis that follows, the most important properties of the space $\mathfrak{F}$ are its completeness and compactness. It might conceivably be possible to define the metric of $\mathfrak{F}$ in some other fashion, such that compactness would result for a greater class of domains than those satisfying condition (A); for such domains and metric, all the results which follow would be valid.

2.4. The customary definitions and conventions are made regarding open and closed sets of elements of $\mathfrak{F}$. In particular the set of all elements of $\mathfrak{F}$ corresponding to boundary elements of T is a closed set, and the set corresponding to points of T is open.

For compact spaces, it is well known that the Heine-Borel theorem holds. If $\mathfrak{F}$ is a closed set of elements of $\mathfrak{F}$ and $\{\Omega\}$ a family of sets with the property that to any element π of $\mathfrak{F}$ there corresponds at least one set Ω such that π is interior to Ω, then there exists a finite set of the sets $\{\Omega\}$ which has the same covering property with respect to $\mathfrak{F}$; that is, $\mathfrak{F}$ is completely interior to $\Omega_1 + \Omega_2 + \Omega_3 + \cdots + \Omega_N$.

Let $\phi(\pi)$ be a function defined on a closed subset $\mathfrak{F}$ of elements of $\mathfrak{F}$. As usual, ϕ is continuous at an element π_0 provided $|\phi(\pi) - \phi(\pi_0)|$ is arbitrarily small with (π, π_0), where π and π_0 are in $\mathfrak{F}$. In terms of points P of T and boundary elements of T, this is what Perkins calls a pseudo-continuous function. In particular $\mathfrak{F}$ may be the set of elements corresponding to all the boundary elements of T; in that case we have a uniformly pseudo-continuous function of boundary elements; we shall henceforth say that such a function is a continuous function of boundary elements.

3. Integral representation. In the case of continuously assigned boundary values $\phi(Q)$, Q on t, and of a domain of which every boundary point is regular, the harmonic function which assumes these boundary values is given by

$$u(M) = \int_t \phi(Q)\, dm(e_Q, M); \tag{2}$$

where $m(e, M)$ is the mass distribution obtained on t by sweeping out a unit mass at M by the Poincaré process onto t.

In this paper we treat first the problem of representing analogously $u(M)$ when $\phi(\gamma)$ is given as a continuous function of the boundary elements. Instead of $m(e, M)$ we shall have a function of sets of boundary elements, which will represent the distribution of the swept out mass according to boundary elements. To this end we shall study general distributions of mass on sets of elements of $\mathfrak{J}$, and sequences of the same in regard to weak convergence. In this connection the reason for making restriction (Λ) becomes evident. For consider the first domain described in § 1. Let $f_n(e)$ be a distribution of mass consisting of unit mass distributed evenly along a line of length $\frac{1}{2}$ lying between the slits at $x = 1/2^n$ and $x = 1/2^{n+1}$. The limiting mass distribution is that of a unit mass spread evenly along the y axis between $y = 0$ and $y = \frac{1}{2}$. Since there are no boundary elements on the segment, no set of boundary elements actually receives any of the limiting mass distribution.

4. A Stieltjes integral. Let $f(\mathcal{E})$ be a positive completely additive function of sets of elements of $\mathfrak{J}$. This may be interpreted, in terms of T and its boundary, as a distribution of positive mass on sets of points of T and boundary elements of T. We wish to set up a linear functional analogous to the Stieltjes integral, and which will be denoted by $\int \phi(\pi) df(\mathcal{E}\pi)$.

Let $\phi(\pi)$ be uniformly continuous on a closed set $\mathfrak{F}$ of $\mathfrak{J}$, and let $f(\mathcal{E})$ be defined and completely additive on a closed family of sets of $\mathfrak{J}$ containing all open and closed sets. By a closed family of sets is meant a family such that all sets obtained from sets of the family by performing finitely or denumerably infinitely many times the operations of addition, multiplication, and taking complements, belong to the family. We suppose also that for any $\mathcal{E}$ on which f is defined, $f(\mathcal{E}) = f(\mathcal{E} \cdot \mathfrak{F})$. Let $\mathfrak{F}$ be divided into finitely or denumerably infinitely many non-overlapping subsets $\mathfrak{F}_i$ belonging to the closed family, and let δ be not less than the diameter of any one of these sets. We can give δ in advance and perform this subdivision; a simple way is to proceed as follows: Construct on the Euclidean space containing T a lattice whose cells have diameter less than δ. This lattice divides T into finitely or denumerably infinitely many domains. To each one of these domains corresponds an open set of elements of $\mathfrak{J}$, and we denote by $\mathfrak{J}_i$ the closed covers of these open sets. Then obviously the diameter of $\mathfrak{J}_i$ is less than δ, and $\mathfrak{F} = \sum_1^\infty \mathfrak{J}_i$; however, the $\mathfrak{J}_i$ are not non-overlapping. We put $\mathfrak{J}'_i = \mathfrak{J}_1$, $\mathfrak{J}'_2 = \mathfrak{J}_2 - \mathfrak{J}_2 \cdot \mathfrak{J}'_1$, $\mathfrak{J}'_3 = \mathfrak{J}_3 - \mathfrak{J}_3 \cdot (\mathfrak{J}'_1 + \mathfrak{J}'_2)$ etc. Then $\mathfrak{F} = \sum_1^\infty \mathfrak{J}'_i$ and

the $\mathcal{J}'_i$ are non-overlapping and belong to the closed family. The sets $\mathcal{J}_i = \mathcal{J} \cdot \mathcal{J}'_i$ will constitute a subdivision of the form required.

Let π_i be any element in $\mathcal{J}_i$ and form the sum

$$S = \sum_1^\infty \phi(\pi_i) f(\mathcal{J}_i).$$

As δ approaches zero, S approaches a unique limit [5] which we denote by

$$\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi).$$

This integral has all the properties analogous to the ordinary Stieltjes integral. In particular, if $\mathcal{J} = \sum_1^\infty \mathcal{J}_i$,

$$\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi) = \sum_1^\infty \int_{\mathcal{J}_i} \phi(\pi)\, df(\mathcal{E}_\pi).$$

If $\phi = \sum_{i=1}^n \phi_i$; $f = \sum_{k=1}^\infty f_k$, then $\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi) = \sum_{i=1}^n \sum_{k=1}^\infty \int_{\mathcal{J}} \phi_i(\pi)\, df_k(\mathcal{E}_\pi)$. If $\lim_{n=\infty} \phi_n = \phi$, uniformly in $\mathcal{J}$, then

$$\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi) = \lim_{n=\infty} \int_{\mathcal{J}} \phi_n(\pi)\, df(\mathcal{E}_\pi).$$

Another useful property is as follows: Let $\omega(\delta)$ be the modulus of continuity of $\phi(\pi)$; that is, $\omega(\delta)$ is the upper bound of $|\phi(\pi) - \phi(\pi')|$ for $(\pi, \pi') \leq \delta$. Then if $\{\mathcal{J}_i\}$ is any subdivision of $\mathcal{J}$ into sets of diameter not greater than δ, and S a sum $\sum_{i=1}^\infty \phi(\pi_i) f(\mathcal{J}_i)$ formed for this subdivision,

$$\left| S - \int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi) \right| \leq \omega(\delta) f(\mathcal{J}).$$

All these properties follow immediately from the definition.

It may be that $f(\mathcal{E})$ represents a mass distribution lying entirely on $\mathcal{J} \subset \mathcal{J}$ and $\phi(\pi)$ is defined only for π on $\mathcal{J}$. Then by $\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi)$ is meant $\int_{\mathcal{J}} \phi(\pi)\, df(\mathcal{E}_\pi)$.

If $\mathcal{J}$ is a set of elements corresponding to a set F of points of T, then

[5] Nelson Dunford, "Integration in General Analysis," *Transactions of the American Mathematical Society*, vol. 37, no. 3, pp. 441-453. An integral is developed in Dunford's paper which includes the one defined here.

ϕ may be interpreted as an ordinary point function on T and f as a function of point sets, and the functional defined above is identical with the Stieltjes integral, $\int_F \phi(P)df(e_\rho)$. Another case of interest is when $\mathfrak{F}$ corresponds to a set Γ of boundary elements; in this case the integral will be denoted by $\int_\Gamma \phi(\gamma)df(\mathcal{E}_\gamma)$. It is this latter form that will subsequently be used to represent the harmonic functions in T.

It may be noted that in defining our generalized Stieltjes integral, we did not use the fact that condition (A) was imposed on T. Thus $\int_{\mathfrak{F}} \phi(\pi)df(\mathcal{E}_\pi)$ may be considered as defined for all domains.

5. Weak convergence of set functions. Consider a sequence $\{f_n(\mathcal{E})\}$ of positive completely additive functions of sets of elements of $\mathfrak{F}$, bounded in their set. It is desired to find a function $f(\mathcal{E})$ and a subsequence $\{\bar{f}_n(\mathcal{E})\}$ such that

$$\lim_{n=\infty} \int_{\mathfrak{F}} \phi(\pi)d\bar{f}_n(\mathcal{E}_n) = \int_{\mathfrak{F}} \phi(\pi)df(\mathcal{E}_\pi),$$

for every continuous $\phi(\pi)$. The sequence of set functions $\{\bar{f}_n(\mathcal{E})\}$ will be said to converge weakly to $f(\mathcal{E})$.

To this end, let there be defined in $\mathfrak{F}$ a denumerably family $\mathfrak{M}$ of open sets I_0 and their closed covers I_f with the following properties:

(a). Given any element π of $\mathfrak{F}$, and a positive number δ, there exists an I_0 (and consequently an I_f) containing π in its interior and having diameter less than δ.

(b). $\mathfrak{F}$ is a member of $\mathfrak{M}$, and any set obtained from sets of $\mathfrak{M}$ by performing finitely many times the operations of addition, multiplication, or taking complements is a member of $\mathfrak{M}$.

That such a family exists is immediate in view of the compactness of $\mathfrak{F}$. For consider for any element π of $\mathfrak{F}$ the set of elements π' such that $(\pi, \pi') < 1$. By the Heine-Borel theorem, a finite number of these cover $\mathfrak{F}$. Select such a finite number, and call their totality, together with their closed covers, $\mathfrak{M}_1$. Similarly consider neighborhoods of elements π of radii $\frac{1}{2}, \frac{1}{4}, \cdots$ and obtain finite families $\mathfrak{M}_2, \mathfrak{M}_3, \cdots$, each covering $\mathfrak{F}$. If to the family $\mathfrak{M}_1 + \mathfrak{M}_2 + \mathfrak{M}_3 + \cdots$ we adjoin $\mathfrak{F}$ and all the sets obtained from this family by performing the operations (b) above, we obtain a family having the required properties.

By a diagonal process, let a subsequence $\{\bar{f}_n(\mathcal{E})\}$ of $\{f_n(\mathcal{E})\}$ be selected such that

$$\lim_{n=\infty} \bar{f}_n(I_0) \quad \text{and} \quad \lim_{n=\infty} \bar{f}_n(I_f)$$

exist for all I_0 and I_f. For convenience in notation, after this is done we shall drop the superscript $^-$ on the functions f and suppose that $\{f_n(\mathcal{E})\}$ is itself the subsequence.

Let $\mathcal{F}$ denote henceforth a closed set of elements of $\mathcal{J}$ and Ω an open set. We define

$$f(\mathcal{F}) = \underset{I_0 \supset \mathcal{F}}{\text{bound}} \{\lim_{n=\infty} f_n(I_0)\};$$

$$f(\Omega) = \overline{\underset{I_f \subset \Omega}{\text{bound}}} \{\lim_{n=\infty} f_n(I_f)\}.$$

$$f(\text{null set}) = 0.$$

THEOREM. *The function $f(\mathcal{E})$ is monotone on the sets on which it is defined.*

If we have $\mathcal{F}_1 \subset \mathcal{F}_2$, $\Omega_1 \subset \Omega_2$, or $\Omega \subset \mathcal{F}$, it is obvious that $f(\mathcal{F}_1) \leq f(\mathcal{F}_2)$, $f(\Omega_1) \leq f(\Omega_2)$, and $f(\Omega) \leq f(\mathcal{F})$ respectively. Consider then $\mathcal{F} \subset \Omega$; we must show that $f(\mathcal{F}) \leq f(\Omega)$. First let it be said that $\delta > 0$ exists such that if π is in $\mathcal{F}$ and π' is in $c\Omega$, $(\pi, \pi') \geq \delta$.

We consider then the aggregate of all sets of the family $\mathfrak{M}$ of type I_0 which have elements in common with $\mathcal{F}$ and which have diameters less than $\delta/2$. Select a finite number of them which cover $\mathcal{F}$, their sum is a set I_0 containing $\mathcal{F}$ with the property that any element contained in it is within a distance $< \delta/2$ of some element of $\mathcal{F}$. In an analogous manner, we can construct a set I_f contained in Ω and having the property that any element of $\Omega - I_f$ is within a distance $< \delta/2$ of some element of $c\Omega$. Then I_f contains I_0. For suppose that I_0 contains an element π not in I_f. Then π' in $\mathcal{F}$ and π'' in $c\Omega$ exist such that $(\pi, \pi') < \delta/2$ and $(\pi, \pi'') < \delta/2$, which is impossible, for then $(\pi', \pi'') < \delta$. We have then

$$f(\mathcal{F}) \leq \lim_{n=\infty} f_n(I_0) \leq \lim_{n=\infty} f_n(I_f) \leq f(\Omega).$$

THEOREM. *On the sets on which it is defined, f is additive in the restricted sense.*

(a). Let $\Omega = \Omega_1 + \Omega_2$, with $\Omega_1 \cdot \Omega_2 = \text{null}$. If I_{f1} and I_{f2} are contained in Ω_1 and Ω_2 respectively, then I_{f1} and I_{f2} have no element in common and $I_{f1} + I_{f2}$ is contained in $\Omega_1 + \Omega_2$. Thus

$$f(\Omega) \geq \lim_{n=\infty} f_n(I_{f1} + I_{f2}) = \lim_{n=\infty} f_n(I_{f1}) + \lim_{n=\infty} f_n(I_{f2}),$$

and since this is true for all I_{f1} and I_{f2},

$$f(\Omega) \geq f(\Omega_1) + f(\Omega_2).$$

Let now I_f be contained in Ω. $I_f \cdot \Omega_1$ and $I_f \cdot \Omega_2$ are non-overlapping and their sum equals I_f. We must show that they are closed. If $I_f \cdot \Omega_1$ is not closed, there must exist an element π, limit of a sequence of elements of $I_f \cdot \Omega_1$, but not in $I_f \cdot \Omega_1$. It must therefore be in Ω_2. Since Ω_2 is open, π contains in its immediate neighborhoods only elements of Ω_2. Consequently $I_f \cdot \Omega_1 = I_{f1}$ and $I_f \cdot \Omega_2 = I_{f2}$ are closed sets and are a positive distance apart.

$$f(\Omega_1) + f(\Omega_1) \geq \lim_{n=\infty} f_n(I_{f1}) + \lim_{n=\infty} f_n(I_{f2}) = \lim_{n=\infty} f_n(I_f),$$

and, since this is true for all I_f,

$$f(\Omega_1) + f(\Omega_2) \geq f(\Omega_1 + \Omega_2).$$

Combining this with the opposite inequality, we have the desired result:

$$f(\Omega_1) + f(\Omega_2) = f(\Omega_1 + \Omega_2).$$

(b). Let $\mathfrak{F} = \mathfrak{F}_1 + \mathfrak{F}_2$, with $\mathfrak{F}_1 \cdot \mathfrak{F}_2 = \text{null}$. By obvious modifications of the method used in proving part (a) of this theorem, one shows that $f(\mathfrak{F}_1) + f(\mathfrak{F}_2) = f(\mathfrak{F}_1 + \mathfrak{F}_2)$; the proof is quite a simple exercise and need not be included here.

If $I_0 \supset \mathfrak{F}$, then $CI_0 \subset C\mathfrak{F}$. Bearing this in mind, it is easy to see that

$$f(\mathfrak{F}) + f(C\mathfrak{F}) = f(\mathfrak{F}).$$

THEOREM. *If* $\Omega = \sum_{i=1}^{\infty} \Omega_i$, *then* $f(\Omega) \leq \sum_{1}^{\infty} f(\Omega_i)$.

Suppose the contrary, that is, for some $\Omega = \sum_{1}^{\infty} \Omega_i$, we have $f(\Omega) > \sum_{i=1}^{\infty} f(\Omega_i)$. Then a closed member of $\mathfrak{M}$, say $\mathfrak{d}_f$ exists such that $\mathfrak{d}_f \subset \Omega$ and

$$\lim_{n=\infty} f_n(\mathfrak{d}_f) > \sum_{1}^{\infty} f(\Omega_i).$$

Now $\mathfrak{d}_f$ is a closed set, each element of which is contained in an Ω_i, and so there exists a finite number of the Ω_i, say the first N, covering $\mathfrak{d}_f$; $\mathfrak{d}_f = \mathfrak{d}_f \cdot \sum_{1}^{N} \Omega_i$. Then a fortiori,

$$\lim_{n=\infty} f_n(\mathfrak{d}_f) > \sum_{i=1}^{N} f(\Omega_i).$$

$$\text{Now } \sum_{i=1}^{N} f(\Omega_i) = \sum_{i=1}^{N} \overline{\underset{I_f \subset \Omega_i}{\text{bound}}} \{\lim_{n=\infty} f_n(I_f)\} \geq \overline{\underset{I_f \subset \sum_1^N \Omega_i}{\text{bound}}} \{\lim_{n=\infty} f_n(I_f)\} \geq \lim_{n=\infty} f_n(\mathfrak{d}_f)$$

which gives a contradiction, yielding the theorem.

For any I_0 containing $\mathfrak{F}$, $f_n(I_0) \geq f_n(\mathfrak{F})$, and so

$$f(\mathfrak{F}) = \underset{\overline{I_0 \supset \mathfrak{F}}}{\text{bound}} \{\lim_{n=\infty} f_n(I_0)\} \geq \overline{\lim_{n=\infty}} f_n(\mathfrak{F}).$$

By the montonity of f, if $\mathfrak{F} \subset \Omega$, $f(\mathfrak{F}) \leq f(\Omega)$, and so

$$f(\Omega) \geq \overline{\underset{\mathfrak{F} \subset \Omega}{\text{bound}}} f(\mathfrak{F}) \geq \overline{\underset{\mathfrak{F} \subset \Omega}{\text{bound}}} \{\overline{\lim_{n=\infty}} f_n(\mathfrak{F})\} \geq \underset{\overline{I_f \subset \Omega}}{\text{bound}} \{\lim_{n=\infty} f_n(I_f)\} = f(\Omega);$$

which gives that

$$f(\Omega) = \overline{\underset{\mathfrak{F} \subset \Omega}{\text{bound}}} f(\mathfrak{F}).$$

Similarly

$$f(\mathfrak{F}) = \underset{\overline{\Omega \supset \mathfrak{F}}}{\text{bound}} f(\Omega).$$

Now let $\mathcal{E}$ be any set of elements of $\mathfrak{F}$. Then there exist open sets containing $\mathcal{E}$ and closed sets contained in $\mathcal{E}$. We define

$$\bar{f}(\mathcal{E}) = \underset{\overline{\Omega \supset \mathcal{E}}}{\text{bound}} f(\Omega);$$

$$\underline{f}(\mathcal{E}) = \overline{\underset{\mathfrak{F} \subset \mathcal{E}}{\text{bound}}} f(\mathfrak{F}).$$

Since for $\mathfrak{F} \subset \mathcal{E} \subset \Omega$, $f(\mathfrak{F}) \leq f(\Omega)$, we see that $\underline{f}(\mathcal{E}) \leq \bar{f}(\mathcal{E})$. In particular, if $\underline{f}(\mathcal{E}) = \bar{f}(\mathcal{E})$, we say that $\mathcal{E}$ is normal with respect to f, and define

$$f(\mathcal{E}) = \bar{f}(\mathcal{E}) = \underline{f}(\mathcal{E}).$$

If $\mathcal{E}$ is open or closed, this definition of $f(\mathcal{E})$ is clearly consistent with the one given previously in terms of I_0 and I_f.

The proof that the function can be extended as a monotone completely additive function may be found in the literature in various places;[6] the three theorems proved above in this section give the sufficient conditions. Among the sets normal with respect to f are all the Borel measurable sets.

THEOREM. *Let $\mathcal{E}$ be a set of elements of $\mathfrak{F}$, normal with respect to f, and let $\mathfrak{G}$ be the closed set of elements constituting the "boundary" of $\mathcal{E}$;*

[6] Hahn, *Theorie der reellen Funktionen*, Berlin, 1921, vol. 1, p. 451. De la Vallée Poussin, *Intégrales de Lebesgue*, Paris, 1934, pp. 23 ff.

that is, π belongs to $\mathfrak{G}$ if within every neighborhood of π there are elements of $\mathcal{E}$ and of $C\mathcal{E}$. Then if $f(\mathfrak{G}) = 0$,

$$f(\mathcal{E}) = \lim_{n=\infty} f_n(\mathcal{E}).$$

Now $\mathcal{E} + \mathfrak{G}$ is a closed set; hence, as has already been proved,

$$f(\mathcal{E} + \mathfrak{G}) \geq \overline{\lim_{n=\infty}} f_n(\mathcal{E} + \mathfrak{G}).$$

Similarly

$$f(\mathcal{E} - \mathfrak{G}) \leq \underline{\lim_{n=\infty}} f_n(\mathcal{E} - \mathfrak{G}).$$

By hypothesis,

$$f(\mathcal{E} + \mathfrak{G}) = f(\mathcal{E} - \mathfrak{G}),$$

and

$$\underline{\lim_{n=\infty}} f_n(\mathcal{E} - \mathfrak{G}) \geq \overline{\lim_{n=\infty}} f_n(\mathcal{E} + \mathfrak{G}).$$

This shows that

$$\underline{\lim_{n=\infty}} f_n(\mathcal{E} - \mathfrak{G}) = \overline{\lim_{n=\infty}} f_n(\mathcal{E} + \mathfrak{G}) = \lim_{n=\infty} f_n(\mathcal{E});$$

or

$$f(\mathcal{E}) = \lim_{n=\infty} f_n(\mathcal{E}).$$

THEOREM. *$\{f_n(\mathcal{E})\}$ converges weakly to $f(\mathcal{E})$.*

Let $\phi(\pi)$ be any continuous function defined on $\mathcal{J}$, $\omega(\delta)$ its modulus of continuity, and M its maximum modulus. Let B be greater than $f(\mathcal{J})$, $f_1(\mathcal{J}), f_2(\mathcal{J}), \cdots$, and let ϵ be any positive number. Choose δ small enough that $\omega(\delta) < \epsilon$. Divide $\mathcal{J}$ into finitely many non-overlapping normal sets $\sum_1^k \mathcal{E}_i$ with the following properties:

(a). $\mathcal{E}_i$ is of diameter $< \delta$.

(b). $f(\mathcal{E}_i) = \lim_{n=\infty} f_n(\mathcal{E}_i)$, $i = 1, 2, \cdots, k$.

To construct such a subdivision, proceed as follows: consider about an element π the neighborhood consisting of all the elements π' such that $(\pi, \pi') < \rho < \delta/2$. Because of the complete additivity of f, only those neighborhoods of π for a denumerable infinitude of values of ρ can have "boundary sets," defined as in the previous theorem, on which f fails to vanish. Consider the rest of the neighborhoods of π, excluding the denumerable infinitude mentioned. Considered for all π, these form a covering family for $\mathcal{J}$, out of which we can pick a finite number, $K_1, K_2, \cdots, K_k$ which cover $\mathcal{J}$. These may be overlapping, but they can be rendered not so. In fact, take

$$\mathcal{E}_1 = K_1; \text{ then } f(\mathcal{E}_1) = \lim_{n=\infty} f_n(\mathcal{E}_1).$$
$$\mathcal{E}_2 = K_2 - K_2K_1.$$

The "boundary" of $\mathcal{E}_2$ consists of certain portions of the "boundaries" of K_1 and of K_2. Since f vanishes on each of these two "boundaries," it also vanishes on the "boundary" of $\mathcal{E}_2$, and

$$f(\mathcal{E}_2) = \lim_{n=\infty} f_n(\mathcal{E}_2).$$

Similarly, putting $\mathcal{E}_3 = K_3 - K_3 \cdot \mathcal{E}_2$, etc., we obtain a non-overlapping family of sets $\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_k$ such that $\mathcal{T} = \sum_1^k \mathcal{E}_i$ and such that the conditions (a) and (b) are satisfied.

With this subdivision,

$$\left| \int_{\mathcal{T}} \phi(\pi)\, df(\mathcal{E}_\pi) - \sum_1^k \phi(\pi_i) f(\mathcal{E}_i) \right| \leq f(\mathcal{T})\omega(\delta) \leq B\epsilon.$$

$$\left| \int_{\mathcal{T}} \phi(\pi)\, df_n(\mathcal{E}_\pi) - \sum_{i=1}^k \phi(\pi'_i) f_n(\mathcal{E}_i) \right| \leq f_n(\mathcal{T})\omega(\delta) \leq B\epsilon.$$

$$\left| \sum_{i=1}^k \phi(\pi_i) f(\mathcal{E}_i) - \sum_{i=1}^k \phi(\pi'_i) f_n(\mathcal{E}_i) \right| \leq M \sum_1^k |f(\mathcal{E}_i) - f_n(\mathcal{E}_i)| + B\epsilon.$$

Since only a finite number k of sets $\mathcal{E}_i$ are involved, we may choose n_0 such that for $n \geq n_0$,

$$\sum_1^k |f(\mathcal{E}_i) - f_n(\mathcal{E}_i)| < \epsilon.$$

Then

$$\left| \int_{\mathcal{T}} \phi(\pi)\, df(\mathcal{E}_\pi) - \int_{\mathcal{T}} \phi(\pi)\, df_n(\mathcal{E}_\pi) \right| < \epsilon(M + 3B),$$

Or

$$\int_{\mathcal{T}} \phi(\pi)\, df(\mathcal{E}_\pi) = \lim_{n=\infty} \int_{\mathcal{T}} \phi(\pi)\, df_n(\mathcal{E}_\pi).$$

That is, $\{f_n(\mathcal{E})\}$ converges weakly to $f(\mathcal{E})$.

6. Generalization of the preceeding. Let τ be a closed set of elements of $\mathcal{T}$ and B a subset thereof. $\theta(\pi)$ is a continuous function defined on τ, and $\phi(\pi)$ is a bounded function defined and continuous in all $\mathcal{T}$ except on B, agreeing with $\theta(\pi)$ on $\tau - B$. The function ϕ need not be defined on B. Let $f(\mathcal{E})$ be a positive set function such that for any $\mathcal{E}$ on which it is defined, $f(\mathcal{E}) = f(\mathcal{E} \cdot \tau - B)$; that is, if f represents a mass distribution on $\mathcal{T}$, all the mass lies on $\tau - B$. We define

$$\int_{\mathfrak{J}} \phi(\pi)\, df(\mathcal{E}_\pi) = \int_{\tau} \theta(\pi)\, df(\mathcal{E}_\pi).$$

This is the natural definition to make, for θ and ϕ differ only on a set on which f is zero.

We consider a sequence $\{f_n(\mathcal{E})\}$ of positive set functions, bounded in their set, with the following property: For any $\epsilon > 0$, there exists an open set $\Omega = \Omega(\epsilon)$ containing B, and an integer $n_0 = n_0(\epsilon)$ such that for $n \geq n_0$, $f_n(\Omega) < \epsilon$. We assume also that the mass of each $f_n(\mathcal{E})$ lies on a closed set of $\mathfrak{J} - \tau$; that is, a closed set $\mathfrak{J}_n$ in $\mathfrak{J} - \tau$ exists such that for any $\mathcal{E}$, $f_n(\mathcal{E}) = f_n(\mathcal{E} \cdot \mathfrak{J}_n)$. Although $\phi(\mathcal{E})$ is not defined on all $\mathfrak{J}$, it is defined on $\mathfrak{J}_n$, and we define

$$\int_{\mathfrak{J}} \phi(\pi)\, df_n(\mathcal{E}_\pi) = \int_{\mathfrak{J}_n} \phi(\pi)\, df_n(\mathcal{E}_\pi).$$

Let $\{\epsilon_i\}$ be a monotone sequence of positive numbers approaching zero, and $\{\Omega_i\}$ the corresponding nested sequence of open sets containing B and such that $f_n(\Omega_i) < \epsilon_i$ for $n \geq n_0(i)$. The set $\mathfrak{J} - \Omega_1$ is a complete and compact space, and $\{f_n(\mathcal{E} \cdot \mathfrak{J} - \Omega_1)\}$ constitutes a sequence of positive set functions thereon. By § 5, we can select a subsequence $\{f_{n_i}(\mathcal{E} \cdot \mathfrak{J} - \Omega_1)\}$ which converges weakly to a positive set function $f^{(1)}(\mathcal{E})$ on $\mathfrak{J} - \Omega_1$. This $f^{(1)}(\mathcal{E})$ may be defined for sets in $\mathfrak{J}$ by $f^{(1)}(\mathcal{E}) = f^{(1)}(\mathcal{E} \cdot \mathfrak{J} - \Omega_1)$. We can do the same thing for $\mathfrak{J} - \Omega_2$, and pick out a subsequence $\{f_{n_{ij}}\}$ of $\{f_{n_i}\}$ such that $\{f_{n_{ij}}(\mathcal{E} \cdot \mathfrak{J} - \Omega_2)\}$, a sequence of set functions on $\mathfrak{J} - \Omega_2$, converges weakly to $f^{(2)}(\mathcal{E})$ on $\mathfrak{J} - \Omega_2$. This can be repeated for each Ω_i, and by a diagonal process, we can select a sequence $\{\bar{f}_n(\mathcal{E})\}$ such that $\{\bar{f}_n(\mathcal{E} \cdot \mathfrak{J} - \Omega_i)\}$ on each $\mathfrak{J} - \Omega_i$ converges weakly to $f^{(i)}(\mathcal{E})$ on $\mathfrak{J} - \Omega_i$. A final selection yields a sequence, which we shall again denote by $\{f_n(\mathcal{E})\}$, which in addition converges weakly to $f(\mathcal{E})$ on $\mathfrak{J}$.

Let us note that for any $\mathcal{E}$, $f^{(i)}(\mathcal{E} \cdot \mathfrak{J} - \Omega_i)$ forms an increasing sequence with i, and this sequence has for limit $f(\mathcal{E})$. This is obvious in view of the fact that for any auxiliary set h of type I_0 or I_f which might be used defining $f(\mathcal{E})$ and the corresponding set $h \cdot \mathfrak{J} - \Omega_i$ which can be used for defining $f^{(i)}(\mathcal{E} \cdot \mathfrak{J} - \Omega_i)$, we have $f_n(h) - f_n(h \cdot \mathfrak{J} - \Omega_i) < \epsilon_i$ for $n \geq n_0(i)$. We see further that

$$f(\mathcal{E}) - f^{(i)}(\mathcal{E} \cdot \mathfrak{J} - \Omega_i) < \epsilon_i.$$

It is evident that $f(B)$ is zero.

THEOREM. $$\int_{\mathfrak{J}} \phi(\pi)\, df(\mathcal{E}_\pi) = \lim_{n=\infty} \int_{\mathfrak{J}} \phi(\pi)\, df_n(\mathcal{E}_\pi).$$

Let M be the upper bound of $\phi(\pi)$. The following inequalities are either evident or immediately verifiable.

$$|\int_{\mathcal{J}} \phi(\pi)\,df(\mathcal{E}_\pi) - \int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df(\mathcal{E}_\pi)| < M\epsilon_i.$$

$$|\int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df(\mathcal{E}_\pi) - \int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df^{(i)}(\mathcal{E}_\pi)| < M\epsilon_i.$$

$$|\int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df_n(\mathcal{E}_\pi) - \int_{\mathcal{J}} \phi(\pi)\,df_n(\mathcal{E}_\pi)| < M\epsilon_i, \quad n \geq n_0(i).$$

Now $\int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df^{(i)}(\mathcal{E}_\pi) = \lim_{n=\infty} \int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df_n(\mathcal{E}_\pi)$, for we may apply the discussion of § 5 to $\mathcal{J}-\Omega_i$, for on this set ϕ is continuous. For sufficiently large n, we have, combining the above inequalities,

$$|\int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df^{(i)}(\mathcal{E}_\pi) - \int_{\mathcal{J}-\Omega_i} \phi(\pi)\,df_n(\mathcal{E}_\pi)| < M\epsilon_i;$$

$$|\int_{\mathcal{J}} \phi(\pi)\,df(\mathcal{E}_\pi) - \int_{\mathcal{J}} \phi(\pi)\,df_n(\mathcal{E}_\pi)| < 4M\epsilon_i.$$

This implies the theorem.

7. Generalization of § 5 to more general domains. Let us consider domains T which do not satisfy condition (A). As suggested earlier, it will be necessary to restrict the sequences of set functions $\{f_n(\mathcal{E})\}$ whose weak convergence we seek to establish. More precisely, we suppose that there exists a sequence of sets $\Omega_1 \supset \Omega_2 \supset \cdots$ of elements of $\mathcal{J}$ and a monotone sequence $\{\epsilon_i\}$ of positive numbers approaching zero, such that

(a). $\mathcal{J}-\Omega_i$ is a complete compact space.

(b). $f_n(\Omega_i) < \epsilon_i$ for $n \geq n_0 = n_0(i)$.

As in § 6, we can select a subsequence, which we denote again by $\{f_n(\mathcal{E})\}$, such that the sequence $\{f_n(\mathcal{E}\cdot\mathcal{J}-\Omega_i)\}$ converge weakly to mass distributions $f^{(i)}(\mathcal{E})$ on $\mathcal{J}-\Omega_i$ for all i. For any $\mathcal{E}$ such that $f^{(i)}(\mathcal{E}\cdot\mathcal{J}-\Omega_i)$ is defined, $\{f^{(i)}(\mathcal{E}\cdot\mathcal{J}-\Omega_i)\}$ is an increasing sequence in i, and we define

$$f(\mathcal{E}) = \lim_{i=\infty} f^{(i)}(\mathcal{E}\cdot\mathcal{J}-\Omega_i).$$

It is readily seen, as in § 6, that

$$f(\mathcal{E}) - f^{(i)}(\mathcal{E}\cdot\mathcal{J}-\Omega_i) < \epsilon_i.$$

The function $f(\mathcal{E})$ is completely additive. First let $\mathcal{E}_1$ and $\mathcal{E}_2$ be non-overlapping.

$$f(\mathcal{E}_1+\mathcal{E}_2)=\lim_{i=\infty} f^{(i)}(\mathcal{E}_1+\mathcal{E}_2\cdot\mathcal{J}-\Omega_i)$$
$$=\lim_{i=\infty} f^{(i)}(\mathcal{E}_1\cdot\mathcal{J}-\Omega_i)+\lim_{i=\infty} f^{(i)}(\mathcal{E}_2\cdot\mathcal{J}-\Omega_i)=f(\mathcal{E}_1)+f(\mathcal{E}_2).$$

Now let $\mathcal{E}=\sum_1^\infty \mathcal{E}_k$, with $\mathcal{E}_i\cdot\mathcal{E}_j=\text{null}$, $i\neq j$. The following are obvious inequalities:

$$0<f(\mathcal{E})-f^{(i)}(\mathcal{E}\cdot\mathcal{J}-\Omega_i)<\epsilon_i,$$
$$0<\sum_{k=1}^n f(\mathcal{E}_k)-\sum_{k=1}^n f^{(i)}(\mathcal{E}_k\cdot\mathcal{J}-\Omega_i)=f(\sum_1^n \mathcal{E}_k)-f^{(i)}(\sum_1^n \mathcal{E}_k\cdot\mathcal{J}-\Omega_i)<\epsilon_i.$$

Allowing n to become infinite,

$$0<\sum_1^\infty f(\mathcal{E}_k)-\sum_1^\infty f^{(i)}(\mathcal{E}_k\cdot\mathcal{J}-\Omega_i)=\sum_1^\infty f(\mathcal{E}_k)-f^{(i)}(\sum_1^\infty \mathcal{E}_k\cdot\mathcal{J}-\Omega_i)<\epsilon_i;$$

for $f^{(i)}(\mathcal{E})$ is known to be completely additive. Consequently,

$$|f(\mathcal{E})-\sum_1^\infty f(\mathcal{E}_k)|<2\epsilon_i,$$

or

$$f(\mathcal{E})=\sum_1^\infty f(\mathcal{E}_k).$$

To show that for $\phi(\pi)$ bounded and uniformly continuous on $\mathcal{J}-\prod_1^\infty \Omega_i$, we have

$$\int_{\mathcal{J}}\phi(\pi)\,df(\mathcal{E}_\pi)=\lim_{n=\infty}\int_{\mathcal{J}}\phi(\pi)\,df_n(\mathcal{E}_\pi),$$

is but to repeat the proof in § 6.

8. Representation of the solution of the Dirichlet problem. Returning to the Dirichlet problem with respect to the domain T, we suppose that T satisfies condition (A) and that all its boundary points are regular.[7] Let $u(\gamma)$ be values assigned continuously to the boundary elements, and $u(P)$ the sequence solution[8] corresponding to these values. Let $\{T_n\}$ be a sequence

[7] A regular point Q of the boundary of T is one at which a barrier can be constructed. A barrier at Q is a function superharmonic in T, zero at Q, and having a positive lower bound outside any sphere about Q. A sufficient condition that Q be a regular point is that a Green's function for T vanish continuously at Q.

[8] Perkins has shown that there exists, corresponding to a continuous $u(\gamma)$, a unique harmonic function $u(P)$, called the sequence solution for the given $u(\gamma)$, with the property that $u(P)$ approaches $u(\gamma)$ as P approaches each boundary element γ located at a regular boundary point.

of nested domains in T with regular boundaries $\{t_n\}$ approximating to t; that is, if F is a closed set of points of T, then $F \cdot T_n$ becomes and remains void for sufficiently large n. M is a point of T, and it will be supposed to be interior to every T. If a unit mass at M is swept out successively onto the boundaries t_n, $n = 1, 2, \cdots$, there is obtained a sequence $\{m_n(e, M)\}$ of positive mass distributions on T. In particular, $m_n(e, M)$ lies entirely on t_n. Furthermore, from (2), we have

$$u(M) = \int_{t_n} u(Q)\, dm_n(e_Q, M).$$

The mass distributions $m_n(e, M)$ may be interpreted, in terms of the notions of § 5, as a sequence of functions of the corresponding sets of elements of $\mathcal{J}$. Denoting them by $\{m_n(\mathcal{E}, M)\}$, we can select a subsequence, which we again denote by $\{m_n(\mathcal{E}, M)\}$, which converges weakly to $m(\mathcal{E}, M)$. In the notation of the abstract space $\mathcal{J}$,

$$u(M) = \int_{\mathcal{J}} u(\pi)\, dm_n(\mathcal{E}_\pi, M) = \lim_{n=\infty} \int_{\mathcal{J}} u(\pi)\, dm_n(\mathcal{E}_\pi, M),$$

and by the weak convergence of $\{m_n(\mathcal{E}, M)\}$,

$$u(M) = \int_{\mathcal{J}} u(\pi)\, dm(\mathcal{E}_\pi, M).$$

It is clear that $m(\mathcal{E}, M)$ represents a mass distribution lying entirely on that closed subset of $\mathcal{J}$ whose elements correspond to boundary elements of T. Consequently, we write the above equation in accordance with the convention set forth in § 4, as the integral over a function of sets of boundary elements.

$$u(M) = \int_t u(\gamma)\, dm(\mathcal{E}_\gamma, M). \tag{3}$$

This gives an integral representation of the solution of the Dirichlet problem with continuously assigned values on the boundary elements, in a regular domain satisfying condition (A).

8. 1. We can also obtain an integral representation for $u(P)$ in domains which have irregular boundary elements but which satisfy condition (A). In that case (3) represents the sequence solution corresponding to the given values.

Let B be the set of irregular points of T. It is known [9] that in the

[9] G. C. Evans, "Potentials of Positive Mass," Part II, *Transactions of the American Mathematical Society*, vol. 38, no. 2, pp. 201-236.

sweeping out process B receives no mass; that is, given $\epsilon > 0$, B can be enclosed in a set of points Ω of $T + t$, open with respect to $T + t$, such that $m_n(\Omega, M) < \epsilon$ for n sufficiently large. Let B' be the set of elements of $\mathcal{J}$ corresponding to irregular boundary points, or in other words corresponding to boundary elements containing points of B. Interpreting the $m_n(\mathcal{E}, M)$ as functions of sets of elements of $\mathcal{J}$, we can enclose B' in an open set Ω' of elements of $\mathcal{J}$ such that for n sufficiently great, $m_n(\Omega', M) < \epsilon$. This is exactly the situation described in § 6. If $u(P)$ is the sequence solution corresponding to the values $u(\gamma)$, it approaches $u(\gamma)$ at all elements except those of B'; hence in accordance with § 6,

$$u(M) = \lim_{n=\infty} \int_{\mathcal{J}} u(\pi)\, dm_n(\mathcal{E}_\pi, M) = \int_{\mathcal{J}} u(\pi)\, dm(\mathcal{E}_\pi, M).$$

Writing this integral in the notation involving sets of boundary elements instead of abstract elements, we have $u(M)$ expressed by (3).

8.2. In extending the representation (3) to certain domains not satisfying condition (A), no attempt will be made to carry through the problem in its fullest generality. A specific example will be considered, and although the method employed therefore will be seen to be applicable to many other domains, it will also be observed that it is ineffective for the most general T.

Consider the plane domain given as the first example of § 1. If we enclose the strip $\mathcal{L}$: $x = 0$, $0 \leqq y < \frac{1}{2}$, in a set Ω, open with respect to the domain, to the remainder of the domain corresponds a portion of the abstract space $\mathcal{J}$ which is compact. Furthermore, if Ω is taken sufficiently small, it will contain as small an amount as desired of the unit mass on the boundaries of the approximating domains used in the sweeping out process. Otherwise the sweeping out process would yield a mass distribution $m(e, M)$ with a positive amount δ deposited on $\mathcal{L}$. Now $m(\mathcal{L}, M)$ is a positive harmonic function [10] of M, never exceeding one, and which vanishes on all boundary points of T except on $\mathcal{L}$. This, however, is impossible,[11] for a conformal representation of T on the interior of the unit circle causes $m(\mathcal{L}, M)$ to be transformed into a bounded harmonic function in the circle vanishing at all points of the circumference except at one point.[12]

We have then precisely the situation treated in § 7, and if $u(M)$ is the sequence solution corresponding to $u(\gamma)$, it is expressed by (3).

---

[10] See footnote 9.

[11] Kellogg, *loc. cit.*

[12] Montel, *Leçons sur les familles normales*, Paris, 1927, pp. 108 ff. The strip $\mathfrak{L}$ constitutes a prime end in the sense of Carathéodory.

The failure of our method of representation to apply except when condition (A) is imposed seems to arise from two causes. Firstly, a treatment of the boundary of T in terms of boundary elements does not adequately take care of all situations that may arise. In the example considered, the strip $\mathfrak{L}$ was an exceptional portion of the boundary. In studying boundaries in general, it is necessary to consider them in terms of Carathéodory's prime ends or the generalization of these prime ends to higher dimensions, but these have not proved to be easily adapted to the problem. Secondly, it is not known that every set of inaccessible prime ends, such as the strip $\mathfrak{L}$ of the boundary of the domain just considered, fails to receive mass in the sweeping out process. In fact, difficulties arise even with simple cases in three dimensions, for there we do not have the conformal mapping as an instrument of proof.

9. Representation of positive harmonic functions. The representation of Maria and Martin for positive harmonic functions in certain domains T is

$$u(P) = \int_t f(S, P)\, d\mu(e), \tag{1}$$

where $f(S, P)$, defined for S on t and P in T, is harmonic in P and continuous in S. It is defined as the Stieltjes derivative $dm(e, P)/dm(e, P_0)$, where $m(e, P)$ is the mass distribution arising from sweeping out unit mass from P onto t. It is equal to one for $P = P_0$, and becomes positively infinite as P approaches S. It vanishes continuously at all other boundary points than S. As an indication of the significance of the formula (1) in terms of familiar ideas, consider a domain T with a boundary sufficiently smooth that the normal derivative of the Green's function exists and is continuous on the boundary, and a harmonic function $u(P)$ assuming continuous boundary values. By applying Green's formula to $u(P)$ and the Green's function $G(P, S)$, we obtain, assuming, for example, that our domain is a three dimensional one,

$$u(P) = -\frac{1}{4\pi}\int_t u(S)\, \frac{\partial G}{\partial \eta_s}(P, S)\, dS.$$

If we put

$$\frac{\partial G}{\partial \eta_s}(P, S) \Big/ \frac{\partial G}{\partial \eta_s}(P_0, S) = f(S, P), \text{ and } \mu(e) = -\frac{1}{4\pi}\int_e u(S)\, \frac{\partial G}{\partial \eta_s}(P_0, S)\, dS,$$

the above equation reduces to (1). In sufficiently smooth domains, then, $f(S, P)$ is merely the normal derivative at S of the Green's function for T with pole at P, multiplied by the proper factor so as to make it equal to one at P_0.

By using the Stieltjes integral for their instrument of representation, Maria and Martin necessarily exclude the domains with boundary points accessible in more than one way; that is, no provision is made for the representation of a harmonic function having different limits as the boundary point is approached in different manners. By employing an analogous integral:

$$u(P) = \int_t f(\gamma, P)\, d\mu(\mathcal{E}_r), \tag{4}$$

with respect to sets of boundary elements, we shall be able to represent the positive harmonic functions in a wider class of domains.

Let T be a bounded domain satisfying condition (A), and let T also satisfy the two following conditions:

CONDITION (B): *Every boundary element of T satisfies a principle of Picard. That is, if $u(P)$ and $v(P)$ are two positive harmonic functions in T which are bounded except in $\mathfrak{S}(\gamma, \rho)$ for all $\rho > 0$, and which vanish at all regular boundary elements of T except γ, then $u(P)/v(P) = \text{constant}$.*

CONDITION (C). *There exists a sequence $T_1 \subset T_2 \subset T \cdots \subset T$ of domains approximating to T, with boundaries $t_1, t_2, \cdots$ and having the properties:*

(a). *T_i admits the Maria-Martin representation of positive harmonic functions; for instance, T is chosen so as to have a continuously turning normal.*

(b). *Let $u_i(P)$ be a positive harmonic function in T_i, zero on t_i except at the point B_i. Let $\{u_i(P)\}$ converge in T to the harmonic function $u(P)$, and let $\{B_i\}$ approach the boundary element γ. Then the $u_i(P)$ are uniformly bounded outside of any $\mathfrak{S}(\gamma, \rho)$.*

Conditions (B) and (C) are natural generalizations of conditions stated by Maria and Martin. We do not, however, include their condition that T be free from irregular boundary points. The three conditions (A), (B), and (C) will be shown to be sufficient that any positive harmonic function in T be given by (4).

The condition (C. b) implies that the function $u(P)$ vanish on all regular boundary elements except γ. For let δ be a regular boundary element other than γ. For ρ sufficiently small, the $u_i(P)$ are uniformly bounded on the portions of their domains of definition that lie within $\mathfrak{S}(\delta, \rho)$. Let $U(P)$ be a barrier at δ; namely, a function positive and superharmonic in T, vanishing at δ, and having a positive lower bound outside any $\mathfrak{S}(\delta, \epsilon)$. That such a function exists at δ is the statement that δ is regular with respect to T. The

$u_i(P)$ are uniformly bounded on the auxiliary boundary of $\mathfrak{S}(\delta, \rho)$, and a suitable positive multiple $kU(P)$ of $U(P)$ will dominate them all on that set of points. Then the function $kU(P) - u(P)$ is superharmonic wherever it is defined in $\mathfrak{S}(\delta, \rho)$, and has no negative boundary value. Hence it is always positive in $\mathfrak{S}(\delta, \rho)$. This shows that near δ, all the $u_i(P)$ are uniformly small, and that $u(P)$ approaches zero as P approaches δ.

For the approximating domain T_n, we may by condition (C. a) define $f_n(S, P)$ such that of $u(P)$ is any positive harmonic function within T, and P is in T_n,

$$u(P) = \int_{t_n} f_n(S, P)\, d\mu_n(e_s), \tag{5}$$

where $\mu_n(e)$ is a distribution of positive mass on t_n. We suppose that P and P_0 are interior to all T_n, and that for all n, $f_n(s, P)$ is chosen so as to be equal to one at P_0. Let γ be any boundary element of T, and S_n a point of t_n such that the sequence $\{S_n\}$ approaches γ. The sequence of positive harmonic functions $\{f_n(S_n, P)\}$ is compact, for they all equal one at P_0. A subsequence $\{\bar{f}_n(\bar{S}_n, P)\}$ converges to a harmonic function which we shall denote by $f(\gamma, P)$. Now $f_n(S_n, P)$ is zero at all points of t_n except S_n; hence by (C. b), the limit function $f(\gamma, P)$, as a function of P, vanishes at all regular boundary elements of T except γ, is bounded except near γ, and becomes positively infinite at γ. Obviously $f(\gamma, P_0) = 1$. If some other subsequence of $\{f_n(S_n, P)\}$ yielded a function $f'(\gamma, P)$, condition (B) would imply that $f'(\gamma, P)$ is identical with $f(\gamma, P)$.

Condition (B) also implies that $f(\gamma, P)$ is continuous in γ. Suppose the contrary; namely that a sequence $\{\gamma_n\}$ exists such that $\lim_{n=\infty} \gamma_n = \gamma$, and that for some P_1,

$$\lim_{n=\infty} f(\gamma_n, P_1) \neq f(\gamma, P_1).$$

Let S_n be chosen on t_n such that $\mathfrak{S}(\gamma_n, 1/n) \supset S_n$, and such that

$$|f_n(S_n, P_1) - f(\gamma_n, P_1)| < 1/n.$$

That this can be done is obvious in view of the manner in which the $f(\gamma_n, P)$ were defined. Of course, there may be required some renumbering of the domains, since subsequences were selected. The sequence $\{S_n\}$ converges to γ, and

$$\lim_{n=\infty} f_n(S_n, P_1) = \lim_{n=\infty} f(\gamma_n, P_1) \neq f(\gamma, P_1).$$

From $\{f_n(S_n, P)\}$ we select a subsequence which converges to a harmonic

function. This function, by condition (B), cannot differ from $f(\gamma, P)$. This contradiction shows that $f(\gamma, P)$ is continuous in γ.

Consider again equation (5). $\{\mu_n(e)\}$ is a sequence of positive mass distributions on closed subsets t_n of T, and since we have from (5) that $\mu_n(t_n) = u(P_0)$, they are all uniformly bounded in total amount. As in § 5 a selection is made of a subsequence, which we again denote by $\{\mu_n(e)\}$, which converges weakly to a mass distribution $\mu(\mathcal{E})$ on the boundary elements. From the manner in which it was defined, it is seen that on the closed set of elements of $\mathcal{T}$ corresponding to the points of the boundaries t_n and the boundary elements of T, the $f_n(S, P)$ and $f(\gamma, P)$ together constitute a continuous function. From the weak convergence, then,

$$\lim_{n=\infty} \int_{t_n} f_n(S, P)\, d\mu_n(e_s) = \int_t f(\gamma, P)\, d\mu(\mathcal{E}_r),$$

or

$$u(P) = \int_t f(\gamma, P)\, d\mu(\mathcal{E}_r).$$

THEOREM. *In a domain T satisfying conditions (A), (B), and (C), any positive harmonic function $u(P)$ is representable in the form* (4).

10. The conditions (B) and (C). One encounters considerable difficulty in attempting to prove the validity of the conditions (B) and (C) for classes of domain of any great degree of generality. Some light may be thrown on the problem, however, by the fact that both are essentially local properties; that is, dependent only on the nature of T in the immediate neighborhood of the particular boundary element named in the condition. A precise formulation of the local character of (B) can be made in the following way: Let $\mathfrak{S}(\gamma, \rho)$ be any pseudo-spherical neighborhood of γ. We consider γ also as a boundary element of the domain $\mathfrak{S}(\gamma, \rho)$, determined by all its pseudo-spherical domains of radius less than ρ. If condition (B) is satisfied for γ with respect to the domain $\mathfrak{S}(\gamma, \rho)$, it is also satisfied for γ with respect to the domain T. The proof of this proposition is exactly analogous to that of Maria and Martin in showing the corresponding fact of a T which is bounded by a finite collection of Jordan curves, and the reader is referred to their paper for its verification.

To make exact the statement that (C) is an "im kleinem" property, consider a $\mathfrak{S}(\gamma, \rho)$ and suppose that the $u_i(P)$ are not uniformly bounded outside $\mathfrak{S}(\gamma, \rho)$, but are uniformly bounded on the auxiliary boundary of $\mathfrak{S}(\gamma, \rho)$. Let M be this supposed uniform bound. Consider any $u_i(P)$ outside $\mathfrak{S}(\gamma, \rho)$. It is defined on the open set of points $T_i(T - \mathfrak{S}(\gamma, \rho))$, which fall into finitely or denumerably infinitely many domains. In any one of these

domains, $u_i(P)$ has no boundary value greater than M. In fact, if P approaches a boundary point of T_i a finite distance from points of the auxiliary boundary of $\mathfrak{S}(\gamma,\rho)$, $u_i(P)$ approaches zero, while on the auxiliary boundary of $\mathfrak{S}(\gamma,\rho)$ $u_i(P)$ assumes values not greater than M. Thus outside $\mathfrak{S}(\gamma,\rho)$ the $u_i(P)$ are uniformly bounded by M, which is contrary to assumption. We can say then that if for the sequence $\{T_n\}$ of approximating domains the condition (C. b) does not hold, then there are boundary elements in every pseudo-spherical neighborhood of γ near which the $u_i(P)$ are not uniformly bounded. To this extent (C) is a local property.

11. A special class of domains. For an example of a class of domains to which we can apply the theory developed in sections one through ten, we consider the class of finitely connected plane regions. No result obtained will be original, for a complete discussion of the representation of harmonic functions in such domains has been given by G. C. Evans,[13] and in fact, his treatment prompted the investigations into the more general cases treated in this paper.

It can be shown that every finitely connected plane region which satisfies (A) also satisfies (B) and (C). In fact, from § 10, we see that we may restrict ourselves to proving (B) and (C) for a $\mathfrak{S}(\gamma,\rho)$, which, in the case of a finitely connected T, either for ρ small enough is simply connected or consists of a circle with center deleted. In the latter case, γ is the boundary element located at an isolated point of t in T, and the proofs of (B) and (C) are trivial. In the former case, we map $\mathfrak{S}(\gamma,\rho) = T'$ conformally on the interior of a circle. This is a 1 : 1 correspondence[14] between the boundary elements of T' and points of the circumference of the circle, and we see that as a result, (B) and (C) are invariant of the conformal mapping. It is known, moreover, that (B) and (C) hold for a circle.[15] They hold also then for T.

We have shown, as an application of the general results obtained in § 9, that in a plane, finitely connected domain satisfying condition (A), every positive harmonic function $u(P)$ is representable in the form (4).

THE UNIVERSITY OF CALIFORNIA,
BERKELEY.

---

[13] See footnote 3.

[14] Montel, *loc. cit.*

[15] Maria and Martin, *loc. cit.*, p. 527.

RIGIDITY FOR SPACES OF CLASS GREATER THAN ONE.*

By Carl B. Allendoerfer.

1. Introduction. In the study of Riemannian geometry there has been considerable interest in those Riemann spaces which are subspaces of Euclidean spaces of one higher dimension, i. e. spaces of class one. Comparatively little is known of spaces of other classes, apparently for the reason that their properties are less intuitively evident. In this paper we consider spaces of class greater than one and define their "type" in an algebraic manner. Thus we are concerned with an open, simply connected domain U_n of a Riemann manifold, R_n, which is a subspace of a Euclidean E_{n+p}.[1] On this basis two theorems are proved, the results of which are valid for any such region within which the hypotheses are fulfilled.

Theorem I. *If the first normal space of U_n is of q $(\leq p)$ dimensions at every point and if U_n is of type ≥ 3 at every point, U_n can be imbedded in an E_{n+q}, and this imbedding is unique to within a rigid motion.*

This theorem is a generalization of Beez's theorem for spaces of class one.[2]

Theorem II. *If the type of $U_n \geq 4$ at every point, the Codazzi equations are consequences of the Gauss equations.* This was proved for spaces of class one by T. Y. Thomas.[3] The proof of Theorem I is contained in § 3 and that of Theorem II is in § 4. A summary of the theory underlying this paper is contained in the author's paper [4] which hereafter will be referred to as "A."

2. The type of a Riemann space. We are concerned with a domain, U_n, of a Riemann manifold; i. e. one whose metric is defined by a number of positive definite quadratic forms $g_{\alpha\beta}$, one in each coördinate neighborhood.

* Received July 8, 1938; Revised October 15, 1938.

[1] This is a definite restriction on R_n. For examples are known of such domains for which no suitable E_{n+p} exists. For instance see L. Bieberbach: "Eine singularitatenfrei Fläche konstanter negativer Krümmung in Hilbertschen Raum," *Commentarii Mathematici Helvetici*, vol, 4 (1932), pp. 248-255. This is true in spite of the classical proof of the existence of such imbeddings in the small.

[2] Compare Eisenhart, *Riemannian Geometry*, § 60. The original paper is: R. Beez: "Zur Theorie der Krümmungsmasses von Mannigfaltigkeiten höherer Ordung," *Zeitschrift für Mathematik und Physik*, vol. 21 (1876), pp. 373-401.

[3] T. Y. Thomas, "Riemann spaces of class one and their characterization," *Acta Mathematica*, vol. 67 (1936), pp. 169-211.

[4] C. B. Allendoerfer, "The imbeddding of Riemann spaces in the large," *Duke Mathematical Journal*, vol. 3, pp. 317-333 (1937).

The coördinates of this space are x^α $(\alpha = 1 \cdots n)$ and Greek indices are reserved for geometric objects belonging to this space.[5] It is a subspace of a Euclidean E_{n+p} whose coördinates are y^i $(i = 1 \cdots n+p)$, Latin indices referring only to this space. We further assume that the imbedding functions $y^i(x)$ are of class C^4 at least. (Many of the results hold even if these are of class C^3). Under these circumstances there exists at each point a "first normal space" of dimension, say, q $(\leq p)$ which consists of the totality of vectors linearly dependent upon the set:

$$Y_{\alpha\beta}{}^i \equiv \frac{\partial^2 y^i}{\partial x^\alpha \partial x^\beta} - \Gamma_{\alpha\beta}{}^\gamma \frac{\partial y^i}{\partial x^\gamma}. \tag{2.1}$$

We assume that in U_n the dimension q remains constant. The question of variable q has not been thoroughly investigated although reference should be made to the work of Wilson and Moore.[6] From among these vectors a mutually orthogonal basis of q unit vectors can be chosen, say $\xi_\sigma{}^i$ $(\sigma = 1 \cdots q)$, and we may write:

$$Y_{\alpha\beta}{}^i = \Omega_{\alpha\beta}{}^\sigma \xi_\sigma{}^i \tag{2.2}$$

where $\Omega_{\alpha\beta}{}^\sigma$ are thus defined as tensors of class C^2. We also define tensors $C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q}$ as follows:

$$C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q} = \begin{vmatrix} \Omega_{\alpha_1\beta_1}{}^1 & \Omega_{\alpha_1\beta_1}{}^2 \cdots \Omega_{\alpha_1\beta_1}{}^q \\ \Omega_{\alpha_2\beta_2}{}^1 & \cdots\cdots \Omega_{\alpha_2\beta_2}{}^q \\ \cdot & \\ \cdot & \\ \cdot & \\ \Omega_{\alpha_q\beta_q}{}^1 & \cdots\cdots \Omega_{\alpha_q\beta_q}{}^q \end{vmatrix}. \tag{2.3}$$

This definition of $C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q}$ is dependent upon the choice of the q-uple $\xi_\sigma{}^i$. For $\Omega_{\alpha\beta}{}^\sigma$ are determined only to within an orthogonal transformation on the index σ. But from the theory of the orthogonal group we have that $C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q}$ is an invariant of the group to within algebraic sign. In what follows we shall be purely concerned with the vanishing and non-vanishing of $C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q}$ which is obviously a property of the space and is not dependent on the choice of $\xi_\sigma{}^i$. Now let:

$$\begin{aligned} C_1 &= C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q} \delta^{\beta_1 \cdots \beta_q}_{k_1 \cdots k_q} \\ C_2 &= C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q} C_{\gamma_1 \ldots \gamma_q | \delta_1 \ldots \delta_q} \delta^{\beta_1 \cdots \beta_q \, \delta_1 \cdots \delta_q}_{k_1 \cdots \quad \cdots k_{2q}} \\ C_3 &= C_{\alpha_1 \ldots \alpha_q | \beta_1 \ldots \beta_q} C_{\gamma_1 \ldots \gamma_q | \delta_1 \ldots \delta_q} C_{\epsilon_1 \ldots \epsilon_q | \zeta_1 \ldots \zeta_q} \delta^{\beta_1 \cdots \beta_q \, \delta_1 \cdots \delta_q \, \zeta_1 \cdots \zeta_q}_{k_1 \cdots \quad \cdots \quad \cdots k_{3q}} \end{aligned} \tag{2.4}$$

[5] The space will, in general be covered by a series of coördinate systems, x^α is meant to be a typical one of these systems.

[6] E. B. Wilson and C. L. E. Moore, "Differential geometry of two-dimensional surfaces in hyperspace," *Proceedings of the American Academy of Arts and Sciences* (Boston), vol. 52 (1916), pp. 269-368.

and similarly for all quantities[7] C_r for $rq \leq n$. Other C_r are defined to be zero. We note that the C_r are symmetric in the α's, in the γ's, in the ϵ's and so forth. Further we note that if $C_r = 0$ at a point, then $C_k = 0$ for $k > r$.

DEFINITION OF TYPE. *If $C_r \neq 0$ and $C_{r+1} = 0$ at a point P, the type of U_n at P is r.*[8] The reader will notice the close connection between this definition and that of the rank of a matrix. The maximum value of the type is thus $[n/q]$. If $q = 1$ (i. e. as in the case of hypersurfaces) the type is the same as the rank of the second fundamental form.

The right-hand sides of (2. 4) are somewhat difficult to handle as they stand, but this complexity may be removed by choosing a suitable coördinate system in which these expressions reduce to determinants at a given point. The existence of such a coördinate system depends on the following lemmas.

LEMMA I. *If a tensor $T_{\alpha \ldots \beta|\gamma \ldots \delta}$ is symmetric in $\alpha \cdots \beta$ and if it is not a zero tensor at a point P, there exists a coördinate system, S, and a set of indices $\mu \cdots \nu$ such that in S the component $T_{1 \ldots 1|\mu \ldots \nu} \neq 0$ at P.*

For

$$\bar{T}_{1 \ldots 1|\mu \ldots \nu} = T_{\alpha \ldots \beta|\gamma \ldots \delta} \frac{\partial x^\alpha}{\partial \bar{x}^1} \cdots \frac{\partial x^\beta}{\partial \bar{x}^1} \frac{\partial x^\gamma}{\partial \bar{x}^\mu} \cdots \frac{\partial x^\delta}{\partial \bar{x}^\nu}.$$

Now if $\bar{T}_{1 \ldots 1|\mu \ldots \nu} = 0$ at P for all choices of $\mu \cdots \nu$ and for all coördinate systems $\bar{x}$, it follows that at P:

$$T_{\alpha \ldots \beta|\gamma \ldots \delta} \frac{\partial x^\alpha}{\partial \bar{x}^1} \cdots \frac{\partial x^\beta}{\partial \bar{x}^1} = 0$$

for all coördinate systems $\bar{x}$. On account of the symmetry in $\alpha \cdots \beta$ this implies that $T_{\alpha \ldots \beta|\gamma \ldots \delta}$ is a zero tensor, which is a contradiction. An immediate extension is

LEMMA II. *If a tensor $T_{\alpha_1 \ldots \beta_1|\alpha_2 \ldots \beta_2| \ldots |\alpha_j \ldots \beta_j|\gamma \ldots \delta}$ $(j \leq n)$ is symmetric in each set of indices $\alpha \cdots \beta$, and if it is not a zero tensor at a point P, there exists a coördinate system, S, and a set of indices $\mu \cdots \nu$ such that in S $T_{1 \ldots 1|2 \ldots 2| \ldots |j \ldots j|\mu \ldots \nu} \neq 0$ at P.*

Applying Lemma II to the expressions in (2. 4) we see that:

LEMMA III. *If U_n is of type r at a point P, there exists a coördinate system, S, and a set of indices $k_1 \cdots k_{rq}$ such that at P*

---

[7] C_r is intended to be a symbol representing the tensor whose components appear on the right of (2. 4). By $C_r = 0$, we mean that C_r is a zero tensor.

[8] Occasionally we shall speak of a set of functions $\Omega_{\alpha\beta}{}^\sigma$ as being of type r when they satisfy this definition, regardless of whether these tensors are connected with a U_n or not.

$$D_r = C_{1\ldots1|\beta_1\ldots\beta_q} C_{2\ldots2|\gamma_1\ldots\gamma_q} \cdots C_{r\ldots r|\delta_1\ldots\delta_q} \delta^{\beta_1\ldots\beta_q\,\gamma_1\ldots\gamma_q\ldots\delta_1\ldots\delta_q}_{k_1\ldots\quad\ldots\quad\ldots\quad\ldots k_{rq}} \neq 0$$

where D_r refers to the particular component expressed on the right.

The importance of this lemma is the fact that the preceeding expression is now Laplace's expansion for a determinant. For $q = 2$ and $r = 3$, it is explicitly

$$D_3 = 8 \begin{vmatrix} \Omega_{1k_1}{}^1 & \Omega_{1k_1}{}^2 & \Omega_{2k_1}{}^1 & \Omega_{2k_2}{}^2 & \Omega_{3k_1}{}^1 & \Omega_{3k_1}{}^2 \\ \Omega_{1k_2}{}^1 & \cdots & \cdots & \cdots & \cdots & \Omega_{3k_2}{}^2 \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ \Omega_{1k_6}{}^1 & \cdots & \cdots & \cdots & \cdots & \Omega_{3k_6}{}^2 \end{vmatrix} . \tag{2.5}$$

Assuming that the type is r, we choose the indices and coördinate system at P as in Lemma III and define $\Omega_\sigma{}^{a\mu}$ as the normalized cofactor of $\Omega_{a\mu}{}^\sigma$ in the non-zero determinant D_r. Then the following relations are true.

$$\Omega_{\mu a}{}^\sigma \Omega_\rho{}^{\nu a} = \delta_\rho{}^\sigma \delta_\mu{}^\nu \tag{2.6}$$

$$\Omega_{\mu a}{}^\sigma \Omega_\rho{}^{\mu a} = r\delta_\rho{}^\sigma \tag{2.7}$$

$$\Omega_{\mu a}{}^\sigma \Omega_\sigma{}^{\mu\beta} = \delta_a{}^\beta. \tag{2.8}$$

$$\begin{pmatrix} \alpha = k_1 \cdots k_{rq} \\ \mu, \nu = 1 \cdots r \\ \sigma, \rho = 1 \cdots q \end{pmatrix}.$$

3. Beez's theorem. In this section we prove Theorem I. First we show that when the type of $U_n \geq 3$, it is possible at imbed U_n in an E_{n+q}. This is a consequence of

THEOREM III. *If the type of $U_n \geq 3$ for every point P, the second curvature tensor vanishes in U_n.* The connection between this theorem and the imbedding problem is discussed in "A." It follows from that paper, equations (5.4) that

$$B_{\xi\eta|\mu a|\lambda\nu} + B_{\xi\eta|\nu a|\mu\lambda} + B_{\xi\eta|\lambda a|\nu\mu} \equiv 0. \tag{3.1}$$

This may be written as

$$\Omega_{\mu a}{}^\sigma \Omega_{\epsilon\theta}{}^\sigma B^{\epsilon\theta}_{\xi\eta\lambda\nu} + \Omega_{\nu a}{}^\sigma \Omega_{\epsilon\theta}{}^\sigma B^{\epsilon\theta}_{\xi\eta\mu\lambda} + \Omega_{\lambda a}{}^\sigma \Omega_{\epsilon\vartheta}{}^\sigma B^{\epsilon\theta}_{\xi\eta\nu\mu} \equiv 0. \tag{3.2}$$

Now at a point P choose the coördinate system, S, described in § 2 in which $D_3 \neq 0$ at P and let the indices have the ranges:

$$\alpha = k_1 \cdots k_{3q};\ \mu, \lambda, \nu = 1, 2, 3;\ \sigma, \rho = 1 \cdots q;\ \epsilon, \theta, \xi, \eta, \pi, \omega = 1 \cdots n.$$

Hence $\Omega_\rho{}^{\mu a}$ may be defined as in § 2. Multiplying (3.2) by $\Omega_\rho{}^{\mu a}$ and summing we find

$$\Omega_{\epsilon\theta}{}^\rho B^{\epsilon\theta}_{\xi\eta\lambda\nu} = 0 \tag{3.3}$$

and consequently

$$\Omega_{\omega\pi}{}^{\rho}\,\Omega_{\epsilon\theta}{}^{\rho}\,B^{\epsilon\theta}_{\xi\eta\lambda\nu} \equiv B_{\omega\pi|\xi\eta|\lambda\nu} = 0 \text{ at } P. \tag{3.4}$$

Replacing ν in (3.2) by an index δ not in the set $1, 2, 3$ and using (3.4) we obtain:

$$\Omega_{\mu a}{}^{\sigma}\,\Omega_{\epsilon\theta}{}^{\sigma}\,B^{\epsilon\theta}_{\xi\eta\lambda\delta} + \Omega_{\lambda a}{}^{\sigma}\,\Omega_{\epsilon\theta}{}^{\sigma}\,B^{\epsilon\theta}_{\xi\eta\delta\mu} = 0. \tag{3.5}$$

Multiplying again by $\Omega_{\rho}{}^{\mu a}$ we have as above that

$$B_{\pi\omega|\xi\eta|\lambda\delta} = 0 \text{ at } P. \tag{3.6}$$

Finally replace λ in (3.5) by κ not in the set $1, 2, 3$ and obtain using (3.6) that:

$$\Omega_{\mu a}{}^{\sigma}\,\Omega_{\epsilon\theta}{}^{\sigma}\,B^{\epsilon\theta}_{\xi\eta\kappa\delta} = 0. \tag{3.7}$$

Multiplying by $\Omega_{\rho}{}^{\mu a}$ and summing as before we see that:

$$B_{\pi\omega|\xi\eta|\kappa\delta} = 0 \text{ at } P. \tag{3.8}$$

Since equations of the form (3.4) hold for all choices of indices, the theorem is proved. This is a sufficient condition that U_n can be imbedded in an E_{n+q}; and it is of prime importance to note that in this case the first normal space is the only normal space. We now suppose this imbedding to have been performed and proceed to show that it is unique to within a rigid motion. To do this we prove Theorems IV and V.

THEOREM IV. *If the type of $U_n \geqq 3$ at a point P, there is only one set of $E_{a\beta|\gamma\delta}$ of rank* [9] *q which satisfy the Gauss equations*

$$B_{a\gamma\delta\beta} = E_{a\beta|\gamma\delta} - E_{a\delta|\gamma\beta} \text{ at } P. \tag{3.9}$$

We recall that by definition

$$E_{a\beta|\gamma\delta} = Y_{a\beta}{}^{i}\,Y_{\gamma\delta}{}^{i}$$

and hence,

$$E_{a\beta|\gamma\delta} = \Omega_{a\beta}{}^{\sigma}\,\Omega_{\gamma\delta}{}^{\sigma}.$$

Let there be two such solutions $E_{a\beta|\gamma\delta}$ and $\bar{E}_{a\beta|\gamma\delta} = \omega_{a\beta}{}^{\sigma}\,\omega_{\gamma\delta}{}^{\sigma}$. Then

$$\Omega_{a\beta}{}^{\sigma}\,\Omega_{\gamma\delta}{}^{\sigma} - \Omega_{a\delta}{}^{\sigma}\,\Omega_{\gamma\beta}{}^{\sigma} = \omega_{a\beta}{}^{\sigma}\,\omega_{\gamma\delta}{}^{\sigma} - \omega_{a\delta}{}^{\sigma}\,\omega_{\gamma\beta}{}^{\sigma}. \tag{3.10}$$

[all indices $1 \cdots n$].

Since the type is at least three, $D_3 \neq 0$ at P in the coördinate system S. Furthermore $\bar{D}_3 \neq 0$, where $\bar{D}_3$ is formed from $\omega_{a\beta}{}^{\sigma}$ in the same manner at D_3 is formed from $\Omega_{a\beta}{}^{\sigma}$. For if $\bar{D}_3 = 0$ at P there exists a non-trivial solution λ^{a} of

[9] The rank of $E_{a\beta|\gamma\delta}$ is that of the matrix whose elements are $E_{a\beta|\gamma\delta}$, $a\beta$ indicating the columns and $\gamma\delta$ the rows. This rank is the same as the dimension of the first normal space.

$$\omega_{\mu\alpha}{}^{\sigma}\,\lambda^{\alpha} = 0 \qquad \begin{pmatrix} \alpha, \beta = k_1 \cdots k_{3q} \\ \mu, \nu = 1, 2, 3 \\ \sigma, \rho = 1 \cdots q \end{pmatrix}. \tag{3.11}$$

However, from (3.10) it follows that

$$\Omega_{\mu\alpha}{}^{\sigma}\,\Omega_{\nu\beta}{}^{\sigma} - \Omega_{\nu\alpha}{}^{\sigma}\,\Omega_{\mu\beta}{}^{\sigma} = \omega_{\mu\alpha}{}^{\sigma}\omega_{\nu\beta}{}^{\sigma} - \omega_{\nu\alpha}{}^{\sigma}\,\omega_{\mu\beta}{}^{\sigma}. \tag{3.12}$$

[indices as in (3.11)]

Multiplying (3.12) by λ^{α} and summing gives as a result of (3.11)

$$(\Omega_{\mu\alpha}{}^{\sigma}\,\Omega_{\nu\beta}{}^{\sigma} - \Omega_{\nu\alpha}{}^{\sigma}\,\Omega_{\mu\beta}{}^{\sigma})\lambda^{\alpha} = 0. \tag{3.13}$$

Now multiply (3.13) by $\Omega_{\rho}{}^{\nu\beta}$ and sum. There results the equation

$$2\Omega_{\mu\alpha}{}^{\rho}\,\lambda^{\alpha} = 0 \tag{3.14}$$

which contradicts the assumption that $D_3 \neq 0$. Thus we may define $\omega_{\sigma}{}^{\mu\alpha}$ in the same manner and for the same range of indices for which $\Omega_{\sigma}{}^{\mu\alpha}$ are defined.

To proceed with the proof of Theorem IV multiply (3.12) by $\Omega_{\rho}{}^{\mu\alpha}$ and sum on μ and α. This gives:

$$2\Omega_{\nu\beta}{}^{\rho} = (\Omega_{\rho}{}^{\mu\alpha}\,\omega_{\mu\alpha}{}^{\sigma})\omega_{\nu\beta}{}^{\sigma} - \Omega_{\rho}{}^{\mu\alpha}\,\omega_{\nu\alpha}{}^{\sigma}\,\omega_{\mu\beta}{}^{\sigma}. \tag{3.15}$$

Again multiply by $\omega_{\tau}{}^{\nu\beta}$ and sum on ν and β, giving:

$$2\omega_{\tau}{}^{\nu\beta}\,\Omega_{\nu\beta}{}^{\rho} = 2\Omega_{\rho}{}^{\mu\alpha}\,\omega_{\mu\alpha}{}^{\tau}. \tag{3.16}$$

Now define new quantities by the equations:

$$\omega_{\tau}{}^{\nu\beta}\,\Omega_{\nu\beta}{}^{\rho} = \Omega_{\rho}{}^{\mu\alpha}\,\omega_{\mu\alpha}{}^{\tau} = 3A_{\tau}{}^{\rho} \tag{3.17}$$

and (3.15) becomes:

$$2\Omega_{\nu\beta}{}^{\rho} = 3A_{\sigma}{}^{\rho}\,\omega_{\nu\beta}{}^{\sigma} - \Omega_{\rho}{}^{\mu\alpha}\,\omega_{\nu\alpha}{}^{\sigma}\,\omega_{\mu\beta}{}^{\sigma}. \tag{3.18}$$

In the same manner we can derive:

$$2\omega_{\nu\alpha}{}^{\sigma} = 3A_{\sigma}{}^{\tau}\,\Omega_{\nu\alpha}{}^{\tau} - \omega_{\sigma}{}^{\lambda\gamma}\,\Omega_{\nu\gamma}{}^{\tau}\,\Omega_{\lambda\alpha}{}^{\tau}. \tag{3.19}$$

Multiplying by $\Omega_{\rho}{}^{\mu\alpha}\,\omega_{\mu\beta}{}^{\sigma}$ and summing on σ and α gives:

$$2\Omega_{\rho}{}^{\mu\alpha}\,\omega_{\mu\beta}{}^{\sigma}\,\omega_{\nu\alpha}{}^{\sigma} = 3A_{\sigma}{}^{\rho}\,\omega_{\nu\beta}{}^{\sigma} - \Omega_{\nu\beta}{}^{\rho}. \tag{3.20}$$

Substituting (3.20) in (3.18) gives:

$$2\Omega_{\nu\beta}{}^{\rho} = 3A_{\sigma}{}^{\rho}\,\omega_{\nu\beta}{}^{\sigma} - \tfrac{3}{2}A_{\sigma}{}^{\rho}\,\omega_{\nu\beta}{}^{\sigma} + \tfrac{1}{2}\Omega_{\nu\beta}{}^{\rho}. \tag{3.21}$$

Therefore it follows that

$$\Omega_{\nu\beta}{}^{\rho} = A_{\sigma}{}^{\rho}\,\omega_{\nu\beta}{}^{\sigma}, \tag{3.22}$$

and similarly

$$\omega_{\nu\beta}{}^{\sigma} = A_{\sigma}{}^{\tau}\,\Omega_{\nu\beta}{}^{\tau}. \tag{3.23}$$

From these equations we see that

$$\Omega_{\nu\beta}{}^{\rho} = A_{\sigma}{}^{\rho}\, A_{\sigma}{}^{\tau}\, \Omega_{\nu\beta}{}^{\tau}; \tag{3.24}$$

or

$$\Omega_{\nu\beta}{}^{\tau}(\delta_{\tau}{}^{\rho} - A_{\sigma}{}^{\rho}\, A_{\sigma}{}^{\tau}) = 0.$$

And multiplying by $\Omega_{\sigma}{}^{\nu\beta}$, we see that

$$A_{\sigma}{}^{\rho}\, A_{\sigma}{}^{\tau} = \delta_{\tau}{}^{\rho} \tag{3.25}$$

i.e. $\| A_{\sigma}{}^{\rho} \|$ is orthogonal. Hence from (3.23) it follows that:

$$E_{a\mu}|_{\beta\nu} = \bar{E}_{a\mu}|_{\beta\nu} \tag{3.26}$$

at every point P. We may extend this result to all sets of indices by the method used in proving Theorem III.

THEOREM V. *If U_n has a first normal space of q dimensions at every point and if with reference to this normal space its type is ≥ 2, U_n cannot be imbedded in any Euclidean space in a manner such that its first normal space is of fewer than q dimensions.*

We have to show that under our hypotheses no solution of the Gauss equations of rank less than q can exist. For at a point P and in the coördinate system S, it is a consequence of (3.9) that

$$B_{12a_1\beta_1}\ldots B_{12a_q\beta_q}\,\delta^{a_1\beta_1\cdots a_q\beta_q}_{k_1\quad\cdots\quad k_{2q}} = 2^q C_{1\ldots1|\beta_1\ldots\beta_q}\, C_{2\ldots2|\gamma_1\ldots\gamma_q}\,\delta^{a_1\beta_1\cdots a_q\beta_q}_{k_1\quad\cdots\quad k_{2q}} \neq 0.^{10} \tag{3.27}$$

Now if a solution of (3.9) of rank $s < q$ existed, say

$$B_{a\gamma\delta\beta} = b_{a\beta}{}^{\sigma}\, b_{\gamma\delta}{}^{\sigma} - b_{a\delta}{}^{\sigma}\, b_{\gamma\beta}{}^{\sigma} \qquad (\sigma = 1\cdots s) \tag{3.28}$$

equations (3.27) would still hold where now

$$C_{1\ldots1|\beta_1\ldots\beta_q} = \begin{vmatrix} b_{1\beta_1}{}^{1}\cdots b_{1\beta_1}{}^{s}\ 0\cdots 0 \\ \cdot \\ \cdot \\ \cdot \\ b_{1\beta_q}{}^{1}\cdots b_{1\beta_q}{}^{s}\ 0\cdots 0 \end{vmatrix}$$ there being $q - s$ columns of zeros.

But this requires the left-hand side of (3.27) to be zero, which is a contradiction.

Since the $g_{a\beta}$ are given, since $E_{a\beta}|_{\gamma\delta}$ of rank q are unique, since no satisfactory $E_{a\beta}|_{\gamma\delta}$ of lower rank exist, and since by Theorem III no other quantities enter into the discussion, the imbedding is unique to within a rigid motion in E_{n+q}. This is the property known as rigidity.[11]

---

[10] The q in 2^q is actually an exponent.

[11] We note that Theorems IV and V do not exclude the possibility of an imbedding for U_n with a first normal space of dimension greater than q.

4. The Codazzi equations. Let us now consider a situation which is the opposite of that heretofore considered. We are now given a U_n and we ask whether it is possible to imbed it in an E_{n+p} with a first normal space of a definite dimension, say q. Necessary and sufficient conditions for this in the small are classical and are known as the Gauss and Codazzi conditions.[12] These require the existence of $\Omega_{\alpha\beta}{}^{\sigma}$ and $T_{\sigma a}{}^{\tau}$ which satisfy the Gauss equations:

$$B_{\alpha\beta\gamma\delta} = \Omega_{\alpha\delta}{}^{\sigma}\,\Omega_{\beta\gamma}{}^{\sigma} - \Omega_{\alpha\gamma}{}^{\sigma}\,\Omega_{\beta\delta}{}^{\sigma} \qquad [\sigma, \tau = 1 \cdots q] \quad [\alpha, \beta, \gamma, \delta = 1 \cdots n] \tag{4.1}$$

and the Codazzi equations:

$$\Omega_{\alpha\beta,\gamma}{}^{\sigma} - \Omega_{\alpha\gamma,\beta}{}^{\sigma} = T_{\sigma\gamma}{}^{\tau}\,\Omega_{\alpha\beta}{}^{\tau} - T_{\sigma\beta}{}^{\tau}\,\Omega_{\alpha\gamma}{}^{\tau} \tag{4.2}$$

and

$$T_{\sigma a,\beta}{}^{\tau} - T_{\sigma\beta,a}{}^{\tau} + T_{\tau a}{}^{\rho}\,T_{\sigma\beta}{}^{\rho} - T_{\tau\beta}{}^{\rho}\,T_{\sigma a}{}^{\rho} + g^{\gamma\delta}(\Omega_{\gamma a}{}^{\tau}\,\Omega_{\delta\beta}{}^{\sigma} - \Omega_{\gamma\beta}{}^{\tau}\,\Omega_{\delta a}{}^{\sigma}) = 0. \tag{4.3}$$

We shall prove

THEOREM II′. *If throughout U_n there exists a set of $\Omega_{\alpha\beta}{}^{\sigma}$ of class C^2 and of type ≥ 4 which satisfy* (4.1), *then there exists in U_n a set of $T_{\sigma a}{}^{\tau}$ of class C^1 which satisfy* (4.2) *and* (4.3). This is a restatement of Theorem II. And in spite of the classical limitation of the Gauss and Codazzi conditions to the small, this theorem is true in the large. A discussion of the imbedding problem in the large has been given by the author in "A," and we treat the application of Theorem II′ to this question in a final paragraph.

To prove Theorem II′ we find by covariant differentiation of (4.1) that

$$B_{\alpha\beta\gamma\delta,\epsilon} = \Omega_{\alpha\delta,\epsilon}{}^{\sigma}\,\Omega_{\beta\gamma}{}^{\sigma} + \Omega_{\alpha\delta}{}^{\sigma}\,\Omega_{\beta\gamma,\epsilon}{}^{\sigma} - \Omega_{\alpha\gamma,\epsilon}{}^{\sigma}\,\Omega_{\beta\delta}{}^{\sigma} - \Omega_{\alpha\gamma}{}^{\sigma}\,\Omega_{\beta\delta,\epsilon}{}^{\sigma}. \tag{4.4}$$

In the same manner it follows that:

$$B_{\alpha\beta\gamma\delta,\epsilon} = \Omega_{\alpha\gamma,\delta}{}^{\sigma}\,\Omega_{\beta\epsilon}{}^{\sigma} + \Omega_{\alpha\gamma}{}^{\sigma}\,\Omega_{\beta\epsilon,\delta}{}^{\sigma} - \Omega_{\alpha\epsilon,\delta}{}^{\sigma}\,\Omega_{\beta\gamma}{}^{\sigma} - \Omega_{\alpha\epsilon}{}^{\sigma}\,\Omega_{\beta\gamma,\delta}{}^{\sigma};$$
$$B_{\alpha\beta\delta\epsilon,\gamma} = \Omega_{\alpha\epsilon,\gamma}{}^{\sigma}\,\Omega_{\beta\delta}{}^{\sigma} + \Omega_{\alpha\epsilon}{}^{\sigma}\,\Omega_{\beta\delta,\gamma}{}^{\sigma} - \Omega_{\alpha\delta,\gamma}{}^{\sigma}\,\Omega_{\beta\epsilon}{}^{\sigma} - \Omega_{\alpha\delta}{}^{\sigma}\,\Omega_{\beta\epsilon,\gamma}{}^{\sigma}.$$

Adding these equations and writing:

$$\phi_{\beta\epsilon\delta}{}^{\sigma} = -\phi_{\beta\delta\epsilon}{}^{\sigma} = \Omega_{\beta\epsilon,\delta}{}^{\sigma} - \Omega_{\beta\delta,\epsilon}{}^{\sigma}, \tag{4.5}$$

we have as a result of Bianchi's identity that:

$$\Omega_{\alpha\gamma}{}^{\sigma}\,\phi_{\beta\delta\epsilon}{}^{\sigma} + \Omega_{\alpha\delta}{}^{\sigma}\,\phi_{\beta\epsilon\gamma}{}^{\sigma} + \Omega_{\alpha\epsilon}{}^{\sigma}\,\phi_{\beta\gamma\delta}{}^{\sigma} + \Omega_{\beta\gamma}{}^{\sigma}\,\phi_{\alpha\epsilon\delta}{}^{\sigma} + \Omega_{\beta\delta}{}^{\sigma}\,\phi_{\alpha\gamma\epsilon}{}^{\sigma} + \Omega_{\beta\epsilon}{}^{\sigma}\,\phi_{\alpha\delta\gamma}{}^{\sigma} = 0. \tag{4.6}$$

Since the type is at least four, we choose the coördinate system, S, and the corresponding set of indices as described in § 2 such that at a point, P, $D_4 \neq 0$. Because of the continuity of the functions involved, $D_4 \neq 0$ in a neighborhood

[12] Compare Eisenhart, *Riemannian Geometry*, p. 190, equations (56.5) to (56.7).

$V(P)$. We now consider the equations (4.6) in $V(P)$ for various ranges of the indices:

Case I. α, β in the range $k_1 \cdots k_{4q}$; γ, δ, ϵ in the range $1, 2, 3, 4$; $\sigma, \rho, \tau = 1 \cdots q$. Multiply (4.6) by $\Omega_\rho{}^{\alpha\gamma}$ and sum on α and γ. This gives:

$$(4.7)\quad 2\phi_{\beta\delta\epsilon}{}^{\rho} + \Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\sigma}\,\phi_{\alpha\epsilon\delta}{}^{\sigma} + \Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\delta}{}^{\sigma}\,\phi_{\alpha\gamma\epsilon}{}^{\sigma} + \Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\epsilon}{}^{\sigma}\,\phi_{\alpha\delta\gamma}{}^{\sigma} = 0.$$

Multiplying (4.7) by $\Omega_\tau{}^{\beta\delta}$ and summing on β and δ we have:

$$(4.8)\quad 2\phi_{\beta\delta\epsilon}{}^{\rho}\,\Omega_\tau{}^{\beta\delta} + 2\Omega_\rho{}^{\alpha\gamma}\,\phi_{\alpha\gamma\epsilon}{}^{\tau} = 0.$$

Now define new quantities $T_{\tau\epsilon}{}^{\rho}$ in $V(P)$ for $\epsilon = 1, 2, 3, 4$ as follows:

$$(4.9)\quad T_{\tau\epsilon}{}^{\rho} = -\,T_{\rho\epsilon}{}^{\tau} = \tfrac{1}{3}\phi_{\alpha\gamma\epsilon}{}^{\tau}\,\Omega_\rho{}^{\alpha\gamma}.$$

We note that these functions are of class C^1 in $V(P)$ since $\Omega_{\alpha\beta}{}^{\sigma}$ are of class C^2. It remains to show that these quantities satisfy (4.2) for suitable choices of the indices. To accomplish this we may write (4.7) in the following form:

$$(4.10)\quad 2\phi_{\beta\delta\epsilon}{}^{\rho} + \Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\sigma}\,\phi_{\alpha\epsilon\delta}{}^{\sigma} - 3\Omega_{\beta\delta}{}^{\sigma}\,T_{\rho\epsilon}{}^{\sigma} + 3\Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\sigma} = 0.$$

Multiplying (4.10) by $\Omega_\tau{}^{\beta\gamma}\,\Omega_{\alpha\gamma}{}^{\rho}$ and summing in β and ρ gives:

$$(4.11)\quad 2\Omega_\tau{}^{\beta\gamma}\,\Omega_{\alpha\gamma}{}^{\rho}\,\phi_{\beta\delta\epsilon}{}^{\rho} + \phi_{\alpha\epsilon\delta}{}^{\tau} - 3\Omega_{\beta\delta}{}^{\rho}\,T_{\rho\epsilon}{}^{\tau} + 3\Omega_{\alpha\epsilon}{}^{\rho}\,T_{\rho\delta}{}^{\tau} = 0.$$

Relettering the indices and using (4.9) we have:

$$(4.12)\quad \phi_{\beta\delta\epsilon}{}^{\rho} + 2\Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\tau}\,\phi_{\alpha\epsilon\delta}{}^{\tau} - 3\Omega_{\beta\delta}{}^{\sigma}\,T_{\rho\epsilon}{}^{\sigma} + 3\Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\sigma} = 0.$$

Subtracting (4.12) from (4.10) gives:

$$(4.13)\quad \phi_{\beta\delta\epsilon}{}^{\rho} - \Omega_\rho{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\tau}\,\phi_{\alpha\epsilon\delta}{}^{\tau} = 0.$$

Hence from (4.10) and (4.13):

$$(4.14)\quad \phi_{\beta\delta\epsilon}{}^{\rho} = \Omega_{\beta\delta}{}^{\sigma}\,T_{\rho\epsilon}{}^{\sigma} - \Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\alpha}.$$

and so the $T_{\rho\epsilon}{}^{\sigma}$ satisfy at least some of (4.2). In order to complete the definition of $T_{\rho\delta}{}^{\sigma}$ when δ is not in the range $1, 2, 3, 4$ we consider

Case II. α, β in the range $k_1 \cdots k_{4q}$; ϵ, γ in the range $1, 2, 3, 4$; δ not in $1, 2, 3, 4$; $\sigma, \rho, \tau = 1 \cdots q$.

Substituting (4.14) in (4.6) we find that:

$$(4.15)\quad \Omega_{\alpha\gamma}{}^{\sigma}\,\bar{\phi}_{\beta\delta\epsilon}{}^{\sigma} - \Omega_{\alpha\epsilon}{}^{\sigma}\,\bar{\phi}_{\beta\delta\gamma}{}^{\sigma} - \Omega_{\beta\gamma}{}^{\sigma}\,\bar{\phi}_{\alpha\delta\epsilon}{}^{\sigma} + \Omega_{\beta\epsilon}{}^{\sigma}\,\bar{\phi}_{\alpha\delta\tau}{}^{\sigma} = 0.$$

where

$$(4.16)\quad \bar{\phi}_{\beta\delta\epsilon}{}^{\sigma} = \phi_{\beta\delta\epsilon}{}^{\sigma} + \Omega_{\beta\delta}{}^{\rho}\,T_{\rho\epsilon}{}^{\sigma}.$$

Multiplying (4.15) by $\Omega_\rho{}^{\alpha\gamma}$ and summing gives:

$$3\bar{\phi}_{\beta\delta\epsilon}{}^{\rho} - \Omega_{\rho}{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\sigma}\,\bar{\phi}_{\alpha\delta\epsilon}{}^{\sigma} + \Omega_{\rho}{}^{\alpha\gamma}\,\Omega_{\beta\epsilon}{}^{\sigma}\,\bar{\phi}_{\alpha\delta\gamma}{}^{\sigma} = 0. \tag{4.17}$$

Multiplying (4.17) by $\Omega_{\tau}{}^{\beta\epsilon}$ and summing gives:

$$3\bar{\phi}_{\beta\delta\epsilon}{}^{\rho}\,\Omega_{\tau}{}^{\beta\epsilon} + 3\Omega_{\rho}{}^{\alpha\gamma}\,\bar{\phi}_{\alpha\delta\gamma}{}^{\tau} = 0. \tag{4.18}$$

So we can define $T_{\tau\delta}{}^{\rho}$ of class C^1 in $V(P)$ (δ not in 1, 2, 3, 4) as follows:

$$T_{\tau\delta}{}^{\rho} = -T_{\rho\delta}{}^{\tau} = \tfrac{1}{4}\Omega_{\tau}{}^{\beta\epsilon}\,\bar{\phi}_{\beta\delta\epsilon}{}^{\rho}. \tag{4.19}$$

To prove that these now satisfy (4.2) for suitable choices of the indices, we observe that (4.17) can now be written:

$$3\bar{\phi}_{\beta\delta\epsilon}{}^{\rho} - \Omega_{\rho}{}^{\alpha\gamma}\,\Omega_{\beta\gamma}{}^{\sigma}\,\bar{\phi}_{\alpha\delta\epsilon}{}^{\sigma} + 4\Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\sigma} = 0. \tag{4.20}$$

And from this it follows as in Case I that

$$\bar{\phi}_{\beta\delta\epsilon}{}^{\rho} = -\Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\sigma}$$

or that

$$\phi_{\beta\delta\epsilon}{}^{\rho} = \Omega_{\beta\delta}{}^{\sigma}\,T_{\rho\epsilon}{}^{\sigma} - \Omega_{\beta\epsilon}{}^{\sigma}\,T_{\rho\delta}{}^{\sigma}. \tag{4.21}$$

Similarly we can show that the $T_{\rho\delta}{}^{\sigma}$ as defined above satisfy (4.2) for all values of the indices. Hence, in general equations of the form (4.2) are satisfied in $V(P)$ as a consequence of the equations (4.1). The new quantities $T_{\rho\delta}{}^{\sigma}$ may now be thought of as the components in S of a scalar in the indices σ and ρ and a tensor in the index δ. And since (4.2) are tensor equations, our result does not depend upon the special coördinate system we have been using.

Since the method of defining $T_{\rho\delta}{}^{\sigma}$ has been quite special, we show that

LEMMA IV. *If $\Omega_{\alpha\beta}{}^{\sigma}$ is of type $\geqq 2$, the $T_{\rho\delta}{}^{\sigma}$ which satisfy* (4.2) (*if any*) *are unique at any point.*

For if both $T_{\rho\delta}{}^{\sigma}$ and $\bar{T}_{\rho\delta}{}^{\sigma}$ satisfy (4.2) we have from those equations that:

$$T_{\sigma\beta}{}^{\tau}\,\Omega_{\alpha\gamma}{}^{\tau} - T_{\sigma\gamma}{}^{\tau}\,\Omega_{\alpha\beta}{}^{\tau} = \bar{T}_{\sigma\beta}{}^{\tau}\,\Omega_{\alpha\gamma}{}^{\tau} - \bar{T}_{\sigma\gamma}{}^{\tau}\,\Omega_{\alpha\beta}{}^{\tau}. \tag{4.22}$$

Again choosing the coördinate system S for which $D_2 \neq 0$ and indices as follows: $\alpha = k_1 \cdots k_{2q}$; $\beta, \gamma = 1, 2$, and multiplying (4.22) by $\Omega_{\rho}{}^{\alpha\gamma}$ and summing we have:

$$T_{\rho\beta}{}^{\tau} = \bar{T}_{\rho\beta}{}^{\tau}. \tag{4.23}$$

This result is easily extended to other values of β by the methods used above. Since we have defined $T_{\rho\beta}{}^{\tau}$ in a unique fashion at each point by (4.9) and (4.19) and since they are of class C^1 in a neighborhood of each such point, it follows that: *$T_{\rho\beta}{}^{\tau}$ as defined are tensors of class C^1 throughout the entire U_n.*

To complete the proof of Theorem II′, we finally demonstrate the

LEMMA V. *When the type of* $\Omega_{\alpha\beta}{}^{\sigma} \geq 3$, *equations* (4.3) *are consequences of* (4.1) *and* (4.2). To prove this we differentiate (4.2) covariantly with respect to δ. Then add to this the two similar equations obtained from it by permuting β, γ, δ cyclicly. The result is:

$$\begin{aligned}(4.24)\quad \Omega_{\alpha\beta,\gamma,\delta}{}^{\sigma} - \Omega_{\alpha\beta,\delta,\gamma}{}^{\sigma} + \Omega_{\alpha\delta,\beta,\gamma}{}^{\sigma} - \Omega_{\alpha\delta,\gamma,\beta}{}^{\sigma} + \Omega_{\alpha\gamma,\delta,\beta}{}^{\sigma} - \Omega_{\alpha\gamma,\beta,\delta}{}^{\sigma} \\ = \Omega_{\alpha\beta}{}^{\tau} F_{\sigma\gamma\delta}{}^{\tau} + \Omega_{\alpha\delta}{}^{\tau} F_{\sigma\beta\gamma}{}^{\tau} + \Omega_{\alpha\gamma}{}^{\tau} F_{\sigma\delta\beta}{}^{\tau};\end{aligned}$$

where

$$(4.25)\qquad F_{\sigma\gamma\delta}{}^{\tau} = T_{\sigma\gamma,\delta}{}^{\tau} - T_{\sigma\delta,\gamma}{}^{\tau} + T_{\tau\gamma}{}^{\rho}\, T_{\sigma\delta}{}^{\rho} - T_{\tau\delta}{}^{\rho}\, T_{\sigma\gamma}{}^{\rho}.$$

But it is well known that:

$$(4.26)\qquad \Omega_{\alpha\beta,\gamma,\delta}{}^{\sigma} - \Omega_{\alpha\beta,\delta,\gamma}{}^{\sigma} = (B_{\theta\beta\delta\gamma}\, \Omega_{\alpha\epsilon}{}^{\sigma} + B_{\theta\alpha\delta\gamma}\, \Omega_{\beta\epsilon}{}^{\sigma})\, g^{\epsilon\theta}.$$

Substituting (4.26) in (4.24) and using (4.1) to simplify the result, we obtain

$$(4.27)\qquad \Omega_{\alpha\beta}{}^{\tau} H_{\sigma\gamma\delta}{}^{\tau} + \Omega_{\alpha\delta}{}^{\tau} H_{\sigma\beta\gamma}{}^{\tau} + \Omega_{\alpha\gamma}{}^{\tau} H_{\sigma\delta\beta}{}^{\tau} = 0,$$

where $H_{\sigma\alpha\beta}{}^{\tau}$ are the left-hand sides of (4.3). Proceeding as in the proof of Theorem III we can show that $H_{\sigma\alpha\beta}{}^{\tau} = 0$ if the type is at least three. This completes the proof of Theorem II′.

5. The imbedding problem in the large. Since a U_n which satisfies the hypotheses of Theorem II′ satisfies those of Theorem I, we can apply the results of the latter and hereafter consider only the case $q = p$. In the author's paper "A," pp. 326-331, it is shown that U_n can be imbedded in E_{n+q} in the large provided that there exists a tensor $E_{\alpha\beta}|_{\gamma\delta}$ $(= \Omega_{\alpha\beta}{}^{\sigma}\, \Omega_{\gamma\delta}{}^{\sigma})$ of class C^1 which satisfies equations (4.1), which satisfies the "Matrix Condition" that there exist $\Gamma^{\mu\nu}_{\alpha\beta\epsilon}$ which satisfy

$$\begin{aligned}(5.1)\quad 2E_{\mu\nu}|_{\gamma\delta}\, \Gamma^{\mu\nu}_{\alpha\beta\epsilon} = 2E_{\mu\nu}|_{\gamma\delta}(\Gamma_{\alpha\epsilon}{}^{\mu}\, \delta_{\beta}{}^{\nu} + \Gamma_{\beta\epsilon}{}^{\mu}\, \delta_{\alpha}{}^{\nu}) + E_{\alpha\beta}|_{\gamma\delta,\epsilon} + E_{\beta\gamma}|_{\delta\epsilon,\alpha} \\ + E_{\gamma\delta}|_{\epsilon\alpha,\beta} - E_{\delta\epsilon}|_{\alpha\beta,\gamma} - E_{\epsilon\alpha}|_{\beta\gamma,\delta},\end{aligned}$$

and whose second curvature tensor vanishes identically. We recall that Theorem III shows the latter to be fulfilled if the type of $E_{\alpha\beta}|_{\gamma\delta}$ is at least three. We now prove:

LEMMA VI. *The "Matrix Condition" is fulfilled for any set of* $E_{\alpha\beta}|_{\gamma\delta}$ *of class* C^1 *satisfying* (4.1) *provided they are of type* ≥ 4.

We prove this by writing $E_{\mu\nu}|_{\gamma\delta} = \Omega_{\mu\nu}{}^{\sigma}\, \Omega_{\gamma\delta}{}^{\sigma}$ throughout (5.1) and

[13] The $\Omega_{\alpha\beta}{}^{\sigma}$ need only be of class C^1 to establish this result. The hypothesis of Theorem II′ which requires $\Omega_{\alpha\beta}{}^{\sigma}$ to be of class C^2 becomes necessary when equations (4.3) are brought into consideration.

using (4. 2), since we know from Theorem II′ that these latter equations are consequences of (4. 1).[13] The result is

$$\Omega_{\gamma\delta^\sigma}\left(\Omega_{\mu\nu^\sigma}\,\Gamma^{\mu\nu}_{\alpha\beta\epsilon}-\frac{\partial\Omega_{\alpha\beta}{}^{\sigma}}{\partial x^{\epsilon}}-\Omega_{\alpha\beta}{}^{\tau}\,T_{\tau\epsilon}{}^{\sigma}\right)=0. \tag{5. 2}$$

Since $\| \Omega_{\gamma\delta^\sigma} \|$ is of rank q, it must follow that:

$$\Omega_{\mu\nu^\sigma}\,\Gamma^{\mu\nu}_{\alpha\beta\epsilon}=\frac{\partial\Omega_{\alpha\beta}{}^{\sigma}}{\partial x^{\epsilon}}+\Omega_{\alpha\beta}{}^{\tau}\,T_{\tau\epsilon}{}^{\sigma}. \tag{5. 3}$$

For any set of $\alpha\beta\epsilon$, these are a set of q linear equations in $n(n+1)/2$ unknowns $\Gamma^{\mu\nu}_{\alpha\beta\epsilon}$. Since the matrix of the coefficients of the unknowns is $\| \Omega_{\mu\nu^\sigma} \|$ which is of rank q, there always exists a solution (not unique) of (5. 3). Consequently the "Matrix Condition" is satisfied. This brings us to the final theorem:

THEOREM VI. *A sufficient condition that a simply connected region, U_n, of a Riemann manifold can be imbedded in a Euclidean E_{n+q} is that there exist a tensor $E_{\alpha\beta}|_{\gamma\delta}$ which satisfies the Gauss equations* (4. 1), *which is of class C^1, which is positive semi-definite of rank q, and whose type is ≥ 4. This imbedding involves only a single* (*i. e. the "first"*) *normal space and is unique to within a rigid motion.*

We note that this is an *algebraic* result, not depending on the solution of any differential equations.

HAVERFORD COLLEGE.

ON A STRETCHING PROCESS FOR SURFACES.*

By T. Radó and P. Reichelderfer.

---

1. In connection with his work on the theory of the area of surfaces, Morrey[1] considers a continuous transformation T operating on the closed region $\mathfrak{R}$ bounded by two concentric spheres having centers at the origin O and radii r and R respectively, where r and R are any positive numbers such that $r < R$. Explicitly his T is given by

$$T: \quad \bar{\mathfrak{x}} = \lambda(\rho)\mathfrak{x}, \quad \lambda(\rho) = (\rho - r)/(R - r), \qquad r \leq \rho \leq R,$$

where $\mathfrak{x}$ and $\bar{\mathfrak{x}}$ represent the vectors extending from O to a point of $\mathfrak{R}$ and to its image under T respectively, and ρ is the length of the vector $\mathfrak{x}$. This T is such that it moves points of $\mathfrak{R}$ along their respective radii vectors drawn from O, so that points on the inner sphere of radius r are thrown into the origin O and points on the outer sphere of radius R remain fixed. Let us consider a smooth surface S in $\mathfrak{R}$—in fact, in Morrey's work S would be a portion of a plane bounded by at most four curves which are either straight lines or arcs of circles. Denote the image of S under T by $\bar{S}$. Let A and $\bar{A}$ be the areas of S and $\bar{S}$ respectively. The argument used by Morrey involves the estimate[2]

$$\bar{A}/A \leqq R/(R - r). \tag{1}$$

The purpose of this note is to present a series of remarks which we arrived at in trying to see more clearly the geometrical background of the inequality (1). Incidentally, our discussion will reveal a slight discrepancy in the statements of Morrey [see § 7].

2. An easy discussion of the applications considered by Morrey would show that any transformation of the form

$$T_\lambda: \quad \bar{\mathfrak{x}} = \lambda(\rho)\mathfrak{x}, \qquad r \leq \rho \leq R,$$

(where $\mathfrak{x}$, $\bar{\mathfrak{x}}$, ρ have meanings as described in § 1) will do, provided only that

---

* Received November 10, 1938; revised December 10, 1938.

[1] C. B. Morrey, "An analytic characterization of surfaces of finite Lebesgue area," part II, *American Journal of Mathematics*, vol. 58 (1936), pp. 312-322. See, in particular, p. 315.

[2] See *loc. cit.*, the last line of p. 315.

the following conditions are satisfied. Firstly, $\lambda(\rho)$ should be chosen from the class K of functions $\lambda(\rho)$ possessing the following properties: i) λ and its derivative λ' are defined and continuous functions of ρ for $r \leq \rho \leq R$; ii) $\lambda(r) = 0$, $\lambda(R) = 1$; iii) $\lambda' > 0$ for $r < \rho < R$. Secondly, an estimate of the form

$$\bar{A}/A \leq F(r, R) \tag{2}$$

should hold for the area magnification of surfaces under T_λ, where $F(r, R)$ is a function such that

$$\liminf F(r, R) \leq 1 \text{ for } r/R \to 0, \qquad R \to 0. \tag{3}$$

On the other hand, it should be emphasized that an estimate of the form (2) is unsatisfactory for the applications if

$$\liminf F(r, R) > 1 \text{ for } r/R \to 0, \qquad R \to 0. \tag{4}$$

3. Given any λ of the class K [see § 2]. Consider the areas A and $\bar{A}$ of a sufficiently smooth surface S in $\mathfrak{R}$ and of its image $\bar{S}$ under T_λ respectively. We want a *sharp* estimate for the area magnification $\bar{A}/A$ for all such surfaces; denote this estimate (the least upper bound of $\bar{A}/A$ for all sufficiently smooth surfaces S in $\mathfrak{R}$) by $U(\lambda)$. Define

$$E(\lambda) = \max_{r \leq \rho \leq R} \frac{1}{2\rho} \frac{d}{d\rho} (\lambda\rho)^2. \tag{5}$$

(6) LEMMA. $$U(\lambda) = E(\lambda).$$

Proof. Let the surface S be given by

$$S: \quad \mathfrak{x} = \mathfrak{x}(u, v), \qquad (u, v) \text{ in } D,$$

where D is some closed bounded region of the uv-plane. Then the area A of S is given by

$$A = \int_D \int W du dv, \text{ where } W^2 = (\mathfrak{x}_u \times \mathfrak{x}_v)^2.$$

The image $\bar{S}$ of S under T_λ is given by

$$\bar{S}: \quad \mathfrak{x} = \bar{\mathfrak{x}}(u, v) = \lambda(\rho)\mathfrak{x}(u, v), \qquad (u, v) \text{ in } D,$$

and its area by

$$\bar{A} = \int_D \int \bar{W} du dv, \text{ where } \bar{W}^2 = (\bar{\mathfrak{x}}_u \times \bar{\mathfrak{x}}_v)^2.$$

Let $M(S,\lambda)$ denote the maximum of $\bar{W}/W$ for (u,v) in D. An easy ϵ-reasoning, left to the reader, shows that $U(\lambda)$ is equal to the least upper bound of $M(S,\lambda)$ for all sufficiently smooth surfaces S in $\mathfrak{R}$. Next, direct computation shows that

$$\bar{W}^2/W^2 = \lambda^2[\lambda^2 + \lambda'\rho(2\lambda + \lambda'\rho)\{1 - (\mathfrak{x}\cdot\zeta/\rho)^2\}], \quad \text{where} \quad \zeta = (\mathfrak{x}_u \times \mathfrak{x}_v)/W.$$

Thus, in view of the conditions on λ [cf. § 2], $\bar{W}/W \leq \lambda(\lambda + \lambda'\rho)$ and hence [cf. formula (5)], $M(S,\lambda) \leq E(\lambda)$ for any of our surfaces S. Consequently $U(\lambda) \leq E(\lambda)$. We want to show that the sign of equality holds in this last relation. Pass a plane p through the origin O; in it choose two radii from O making an angle ϕ less than 180°. Denote the smaller portion of the plane p in $\mathfrak{R}$ bounded by these two radii by S_0. At every point of S_0, $\mathfrak{x}\cdot\zeta = 0$, so $\bar{W}/W = \lambda(\lambda + \lambda'\rho)$. Thus $U(\lambda) \geq M(S_0,\lambda) = E(\lambda)$, and the lemma is established.

4. We next seek a λ in the class K [cf. § 2] for which the estimate $E(\lambda)$ [cf. formula (5)] for the area magnification $\bar{A}/A$ of sufficiently smooth surfaces S in $\mathfrak{R}$ under T_λ is as small as possible. The greatest lower bound of $E(\lambda)$ for all λ in K is a quantity depending only upon r and R; denote it by $L(r,R)$. In view of the remarks in § 2, the actual determination of $L(r,R)$ appears to be a question of some interest.

THEOREM. *We have* $L(r,R) = R^2/(R^2 - r^2)$, *but there does not exist in the class* K *any function* $\lambda(\rho)$ *such that* $E(\lambda) = R^2/(R^2 - r^2)$—*that is,* $R^2/(R^2 - r^2)$ *is the greatest lower bound, but not the minimum of* $E(\lambda)$ *for* λ *in* K.

This theorem is implied by the following three facts which we shall state and prove presently: 1) for every function of the class K [cf. § 2], we have $E(\lambda) \geq R^2/(R^2 - r^2)$; 2) for no λ in K does $E(\lambda) = R^2/(R^2 - r^2)$; 3) for every $\epsilon > 0$ there exists a λ in K such that $E(\lambda) < (1+\epsilon)R^2/(R^2 - r^2)$.

Proofs. 1) Consider the surface S_0 defined in § 3. The area A_0 of S_0 is $(R^2 - r^2)\phi/2$. Any transformation T_λ, λ in K, transforms S_0 into a sector $\bar{S}_0$ of the circle in p with center O and radius R bounded by the same two radii. The area $\bar{A}_0$ of $\bar{S}_0$ is $R^2\phi/2$. Hence [cf. formula (6)],

$$E(\lambda) = U(\lambda) \geq \bar{A}_0/A_0 = R^2/(R^2 - r^2).$$

Since this is valid for any λ in K, we have

$$L(r,R) \geq R^2/(R^2 - r^2).$$

2) Assume there exists a $\lambda = \lambda^{(1)}$ in K such that

$$R^2/(R^2 - r^2) = E(\lambda^{(1)}). \tag{7}$$

Thus

$$\frac{d}{d\rho}(\lambda^{(1)}\rho) \leq 2\frac{R^2}{R^2 - r^2}\rho. \tag{8}$$

Integrating (8) with respect to ρ from r to R, and using condition ii) of § 2, we obtain $R^2 \leq R^2$. Thus the sign of equality must hold in (7); that is,

$$\frac{d}{d\rho}(\lambda^{(1)}\rho) = 2\frac{R^2}{R^2 - r^2}\rho. \tag{9}$$

Integrating (9) with respect to ρ from r to ρ and solving for $\lambda^{(1)}$, we get

$$\lambda^{(1)}(\rho) = \frac{R}{(R^2 - r^2)^{\frac{1}{2}}}\frac{(\rho^2 - r^2)^{\frac{1}{2}}}{\rho}, \qquad r \leqq \rho \leqq R. \tag{10}$$

Hence

$$\frac{d}{d\rho}\lambda^{(1)} = \frac{Rr^2}{(R^2 - r^2)^{\frac{1}{2}}}\frac{1}{\rho^2(\rho^2 - r^2)^{\frac{1}{2}}}.$$

Thus $\lambda'^{(1)}$ does not exist for $\rho = r$, and so $\lambda^{(1)}$ is not in our class K. In fact, since $\lambda'^{(1)} \to +\infty$ for $\rho \to r$, we proved that the relation (7) cannot hold even if we relax the condition that λ' be continuous for $r \leq \rho \leq R$ to the condition that λ' be bounded for $r < \rho < R$. 3) Pick θ so that $0 < \theta \leq 1$, set $\xi = r + \theta(R - r)$, and

$$\lambda_\theta(\rho) = \begin{cases} \dfrac{\xi^{\frac{1}{2}}}{(\xi - r)^{\frac{1}{2}}}\dfrac{R}{(R^2 - \xi r)^{\frac{1}{2}}}\dfrac{\rho - r}{\rho}, & r \leq \rho \leq \xi, \\ \dfrac{R}{(R^2 - \xi r)^{\frac{1}{2}}}\dfrac{(\rho^2 - \xi r)^{\frac{1}{2}}}{\rho}, & \xi \leq \rho \leq R. \end{cases} \tag{11}$$

It is easily verified that, for any θ, $0 < \theta \leq 1$, λ_θ is in the class K. Writing, for greater clarity, $E(\lambda_\theta) = F_\theta(r, R)$ we find by formula (5)

$$F_\theta(r, R) = \frac{R^2}{(R - r)\{R + r(1 - \theta)\}}. \tag{12}$$

Given $\epsilon > 0$, choose θ_ϵ so that $0 < \theta_\epsilon \leq 1$, $\theta_\epsilon < \epsilon(r + R)/r(1 + \epsilon)$; then $E(\lambda_{\theta\epsilon}) < (1 + \epsilon)R^2/(R^2 - r^2)$, as asserted.

5. For θ sufficiently near 0, formulas (11) yield a function $\lambda_\theta(\rho)$ in K for which the estimate $E(\lambda_\theta)$ for the area magnification $\bar{A}/A$ of surfaces S in $\mathfrak{R}$ is arbitrarily close to the lower bound $L(r, R) = R^2/(R^2 - r^2)$, [cf. formula (12)]. For brevity, we do not discuss the somewhat elaborate train

of thought which led us to formulas (11). We just observe that a more obvious procedure will not lead to a λ in K for which $E(\lambda)$ is close to $R^2/(R^2-r^2)$. In § 4, formula (10), we have a function $\lambda^{(1)}$ for which $E(\lambda^{(1)}) = R^2/(R^2-r^2) = L(r,R)$. Hence $\lambda^{(1)}$ would be the best possible function for our purposes, except that it is not in K solely because $\lambda'^{(1)}$ does not behave properly for $\rho = r$ [cf. § 4]. This observation suggests that we modify $\lambda^{(1)}$ in a short interval $r \leq \rho \leq \xi$, by replacing it there by a linear function of ρ. Thus we are led to consider

$$\lambda^{(2)}(\rho) = \begin{cases} \{(\rho-r)/(\xi-r)\}\lambda^{(1)}(\xi), & r \leq \rho \leq \xi, \\ \lambda^{(1)}(\rho), & \xi \leq \rho \leq R. \end{cases} \tag{13}$$

This function satisfies all the conditions of § 2 except that $\lambda'^{(2)}$ does not exist for $\rho = \xi$; this situation may be remedied by smoothing the function in a neighborhood of $\rho = \xi$ so as to alter $E(\lambda)$ but slightly. We have by formula (5), $E(\lambda^{(2)}) = (2 + r/\xi - r^2/\xi^2)R^2/(R^2-r^2) > 2R^2/(R^2-r^2)$ for every choice of r, R, and ξ, $0 < r < \xi < R$. Thus this slight modification of $\lambda^{(1)}$, described by formulas (13), not only more than doubles the corresponding value of $E(\lambda)$, but also yields a function $\lambda^{(2)}$ which will not do for the applications [cf. § 2].

6. From formula (12) it is clear that $\lim F_\theta(r,R) = 1$ for $r/R \to 0$, $R \to 0$, and for every choice of θ, $0 < \theta \leq 1$. Thus every one of the functions $\lambda_\theta(\rho)$ will do for the applications [cf. § 2]. The closer we choose θ to 0 the better the estimate [cf. formula (12)] we obtain for the area magnification $\bar{A}/A$ [cf. formula (6)]. Thus, in a sense, the least favorable of our functions $\lambda_\theta(\rho)$ is obtained for $\theta = 1$—viz.,

$$\lambda_1(\rho) = \frac{R}{R-r}\,\frac{\rho-r}{\rho}, \qquad r \leq \rho \leq R. \tag{14}$$

On the other hand, the function (14) has the simplest expression of all the functions $\lambda_\theta(\rho)$, $0 < \theta \leq 1$.

7. As stated in § 1 of this note, Morrey [loc. cit.[1]] uses the function $\lambda(\rho) = (\rho-r)/(R-r)$, which leads to the estimate [cf. formula (6)]

$$\bar{A}/A \leq 2 + r/(R-r), \tag{15}$$

not to the estimate (1) which is involved in the work of Morrey. Since $2 + r/(R-r) \to 2$ for $r/R \to 0$, $R \to 0$, the estimate (15) is unsatisfactory for the applications [cf. § 2]. Professor Morrey, to whom the manuscript of this note was submitted for comments, kindly gave us the following explana-

tion of the source of this discrepancy. In the paper of Professor Morrey, vector notation is used throughout; accordingly, the transformation T_λ appears (in our notations) in the form $\bar{\mathfrak{x}} = \lambda(\rho)\mathfrak{x}$. However, the same transformation can be described by a single scalar equation of the form $\bar{\rho} = \mu(\rho)$, where ρ and $\bar{\rho}$ are the lengths of $\mathfrak{x}$ and $\bar{\mathfrak{x}}$ respectively. Corresponding functions $\lambda(\rho)$ and $\mu(\rho)$ are related by the equation $\mu(\rho) = \rho\lambda(\rho)$. This latter form of the transformation is the one which Professor Morrey said he originally had in mind. If the transformation be written in the form $\bar{\rho} = \mu(\rho)$, it is quite natural to choose

$$\mu(\rho) = \frac{R(\rho - r)}{R - r}. \tag{16}$$

The corresponding function $\lambda(\rho)$ has the expression

$$\lambda(\rho) = \frac{R(\rho - r)}{\rho(R - r)}. \tag{17}$$

For this function, the estimate

$$\bar{A}/A \leq R/(R - r) \tag{18}$$

actually holds.—For the reader who desires to verify this last statement, we remark that the function (17) is identical with the function denoted by $\lambda_1(\rho)$ in § 6. Therefore the inequality (18) follows from the general inequality (12) for $\theta = 1$.

8. We discussed in this note what may be called a process for *stretching a surface toward a point.* The applications suggest the interest of a process for *stretching a surface toward a line.* We hope to discuss this question on another occasion.

THE OHIO STATE UNIVERSITY.

ON THE PROBLEM OF TEMPERATURES IN A NON-HOMOGENEOUS BAR WITH DISCONTINUOUS INITIAL TEMPERATURES.*†

By R. V. CHURCHILL.

1. Introduction. We are concerned here with the solution of the heat equation in the form

$$U_{xx}(x,t) - q(x)U(x,t) = U_t(x,t) \qquad (0 < x < 1,\ t > 0), \tag{1}$$

where the subscripts denote partial derivatives, under the boundary conditions

$$U(x,+0) = F(x) \qquad (0 < x < 1), \tag{2}$$

$$U(+0,t) = 0, \qquad U(1-0,t) = 0 \qquad (t > 0). \tag{3}$$

It will be assumed throughout that $q(x)$ is continuous in the interval $0 \leqq x \leqq 1$, and that this interval can be divided into a finite number of closed subintervals in each of which $F(x)$ is continuous; that is, $F(x)$ is piecewise continuous in $0 \leqq x \leqq 1$. $F'(x)$ is assumed to be bounded and integrable in that interval.

We are particularly concerned with the determination of a unique solution of this problem and the asymptotic properties of this solution.

First we shall establish a solution and demonstrate at the same time that it can be written in either of two forms: a series form and an integral form. To determine the behavior of the temperature function and its derivatives for large values of t the series form is advantageous, but for small values of t the integral form is much more useful.

A uniqueness theorem will then be established. Following this, additional conditions on the derivatives of $F(x)$ and $q(x)$ will be given under which our solution is the only possible one satisfying certain natural restrictions.

The series form given here is the classical one. It can easily be established as a solution with the aid of a theorem for the expansion of $F(x)$ in a series of the characteristic functions of the corresponding Sturm-Liouville system. The expansion theorem introduces some additional restrictions however. It is not used in the present method.

But the serious defect in the classical treatment is that, when $F(x)$ is discontinuous or fails to satisfy either end condition, the uniqueness of the

* Received August 5, 1938; Revised December 7, 1938.

† Presented to the American Mathematical Society, September 6, 1938.

solution has not been established.[1] Since restrictions on the temperature function and its derivatives near $t = 0$ are essential for uniqueness, serious difficulties can be expected in an attempt to establish uniqueness by using the series form alone.

We shall make use of the Laplace transform of the temperature function,

$$L\{U\} = \int_0^\infty e^{-st} U(x, t)\, dt = u(x, s).$$

This satisfies, at least formally, a boundary value problem in ordinary differential equations containing the complex parameter s. Letting $u(x, s)$ represent the solution of this problem, we shall show that the inversion integral

$$L^{-1}_{(i)}\{u\} = \frac{1}{2\pi i} \lim_{\omega \to \infty} \int_{\gamma - \omega i}^{\gamma + \omega i} e^{tz} u(x, z)\, dz, \tag{4}$$

which can be written as a real infinite integral, represents a solution of the original problem and has $u(x, s)$ for its transform when $\Re(s) > \gamma$. Several properties of the temperature function and its derivatives will be determined from the character of the transforms of those functions.

It should be noted that it is not an essential generalization to replace the heat equation (1) by the equation

$$(k(x) U_x)_x - l(x) U = g(x) U_t$$

when k, g, l are continuous, k and g positive, and k' and $(kg)''$ are continuous for $0 \leqq x \leqq 1$. A substitution of new variables reduces this to (1).[2]

2. The inversion integral. When the function (4) does not vanish as t approaches zero, its derivative with respect to t can not be found by simply differentiating inside the integral. The integral forms of the derivatives, and their limits at $t = 0$, are given by the following

LEMMA 1. *In some half-plane $\Re(s) \geqq \gamma > 0$ let $\theta(s)$ be analytic and have the form*

$$\theta(s) = \frac{A_0}{s} + \frac{A_1}{s^2} + \cdots + \frac{A_{n-1}}{s^n} + O\left(\frac{1}{|s|^{n+\delta}}\right),$$

where n is a positive integer, $A_0, A_1, \cdots, A_{n-1}$ and δ are constants, and $\delta > 0$. Then the inverse transform of $\theta(s)$ is

---

[1] See G. Doetsch, "Les équations aux dérivées partielles du type parabolique," *L'Enseignement Mathématique*, vol. 35 (1936), pp. 43-87, for a discussion of known uniqueness theorems, and examples of multiple solutions.

[2] See for instance the corresponding substitution for an ordinary differential equation in Ince, *Ordinary Differential Equations* (1927), p. 270.

$$\Theta(t) = L_{(i)}^{-1}\{\theta\},$$

that is, $L\{\Theta\} = \theta(s)$ *if* $\Re(s) > \gamma$. *Moreover* $\Theta(t)$ *and its first* $n-1$ *derivatives are continuous for each finite* $t > 0$ *and, for the same* γ,

$$\Theta^{(k)}(t) = L_{(i)}^{-1}\{s^k\theta(s) - A_0 s^{k-1} - A_1 s^{k-2} - \cdots - A_{k-1}\} \qquad (t > 0)$$

and

$$\Theta(+0) = A_0, \qquad \Theta^{(k)}(+0) = A_k \qquad (k = 1, 2, \cdots, n-1).$$

The inversion integral of s^{-1} converges uniformly to 1 in every finite interval in $t > 0$. Also $L_{(i)}^{-1}\{O(|s|^{-1-\delta})\}$ converges uniformly for all $t \geqq 0$; it is known that this vanishes as t approaches zero and consequently that $L_{(i)}^{-1}\{\theta\}$ converges for $t > 0$ to the continuous inverse transform of $\theta(s)$ and has the limit A_0 as t approaches zero.[3]

Let us write

$$\phi(s) = \theta(s) - \frac{A_0}{s} - \frac{A_1}{s^2} - \cdots - \frac{A_{k-1}}{s^k} \qquad (k = 1, 2, \cdots, n-1),$$

then

$$s^k\phi(s) = \frac{A_k}{s} + O\left(\frac{1}{|s|^{1+\delta}}\right).$$

According to a known theorem [4] the function $\Phi(t) = L_{(i)}^{-1}\{\phi\}$ is therefore the inverse transform of $\phi(s)$, and

$$\Phi^{(k)}(t) = L_{(i)}^{-1}\{s^k\phi(s)\};$$

also this is continuous for each finite $t > 0$. Our expression for $s^k\phi(s)$ shows that $\Phi^{(k)}(+0) = A_k$. Since $L_{(i)}^{-1}\{A_m s^{-m-1}\} = A_m t^m/m!$ if $t > 0$,

$$\Phi(t) = \Theta(t) - A_0 - A_1 t - \cdots - A_{k-1}t^{k-1}/(k-1)!.$$

Hence $\Phi^{(k)}(t) = \Theta^{(k)}(t)$, and the lemma is established.

3. The transformed problem. According to the property [5]

$$L\{U_t\} = sL\{U\} - F(x),$$

the problem in the transform $u(x, s)$ corresponding to the problem (1)–(3) is

$$u_{xx}(x, s) - (q(x) + s)u(x, s) = -F(x), \tag{5}$$

$$u(0, s) = 0, \qquad u(1, s) = 0. \tag{6}$$

---

[3] R. V. Churchill, "The solution of linear boundary value problems in physics by means of the Laplace transformation," Part I, *Mathematische Annalen*, vol. 114 (1937), pp. 591-613; Theorems 3 and 4.

[4] Theorem 6 of the paper cited under 3.

[5] G. Doetsch, *Theorie und Anwendung der Laplace-transformation* (1937), p. 152.

This system is found by formally applying the operator L to (1) and (3) and interchanging the order of that operation with differentiation and limits with respect to x. For the present it is not necessary to consider the conditions on $U(x, t)$ and the region of the complex parameter s under which the problem (5), (6) is actually a consequence of the original.

Henceforth we let $u(x, s)$ denote the following solution of the ordinary differential equation (5) satisfying the conditions (6):

$$u(x, s) = \int_0^1 G(x, \xi, s) F(\xi) d\xi, \tag{7}$$

where the Green's function G can be written

$$\begin{aligned} G(x, \xi, s) &= -u_1(x, s) u_2(\xi, s)/D(s) \text{ if } x \leqq \xi, \\ &= -u_2(x, s) u_1(\xi, s)/D(s) \text{ if } x \geqq \xi; \end{aligned} \tag{8}$$

$$D(s) = u_1(1, s). \tag{9}$$

The functions u_1 and u_2 are the unique solutions of the homogeneous equation

$$u_{xx} - (q + s)u = 0 \tag{10}$$

which, together with their first derivatives with respect to x, are continuous in x $(0 \leqq x \leqq 1)$ and satisfy the conditions

$$u_1(0, s) = 0, \quad \frac{du_1(0, s)}{dx} = 1; \qquad u_2(1, s) = 0, \quad \frac{du_2(1, s)}{dx} = 1.$$

These functions $u_1(x, s)$ and $u_2(x, s)$ and their first derivatives are analytic in s and continuous in (x, s) $(0 \leqq x \leqq 1)$ for every s in the finite complex plane.[6] Since $F(x)$ is piecewise continuous it follows readily from (7) and (8) that u and u_x are also analytic in s and continuous in (x, s), when $0 \leqq x \leqq 1$ and s is finite and not a zero of $D(s)$. Also, u_{xx} has these properties when x is not a point of discontinuity of $F(x)$.

The solutions u_1, u_2 of (10) satisfy the equations

$$u_1(x, s)\sqrt{s} = sh\, x\sqrt{s} + \int_0^x u_1(\xi, s) q(\xi) sh(x - \xi)\sqrt{s}\, d\xi, \tag{11}$$

$$u_2(x, s)\sqrt{s} = -sh(1 - x)\sqrt{s} + \int_x^1 u_2(\xi, s) q(\xi) sh(\xi - x)\sqrt{s}\, d\xi. \tag{12}$$

Since u_1, u_2 are analytic in s we need only to be consistent in our use of a branch of $\sqrt{s}$. When $s = \rho e^{i\theta}$ we take $-\pi < \theta \leqq \pi$; so $\Re(\sqrt{s}) \geqq 0$.

[6] The method of establishing their regularity in s by successive approximations, as presented for instance on p. 72 of the book cited under 2, also shows the two-dimensional continuity in (x, s).

For any fixed s let $M(s)$ be the maximum value of $|\, u_1(x,s)\sqrt{s}\, e^{-x\sqrt{s}}\,|$ for all x $(0 \leqq x \leqq 1)$. Since u_1 is continuous in x that value is assumed for some x and it follows from (11) that

$$M(s) \leqq |\, e^{-x\sqrt{s}} sh\, x\sqrt{s}\,| + \rho^{-\frac{1}{2}} \int_0^x |\, u_1(\xi,s)\sqrt{s}\, e^{-\xi\sqrt{s}} q(\xi) e^{-(x-\xi)\sqrt{s}} sh(x-\xi)\sqrt{s}\,|\, d\xi$$

$$\leqq 1 + \rho^{-\frac{1}{2}} M(s) \int_0^1 |\, q(\xi)\,|\, d\xi.$$

Hence for some $\rho_0 > 0$,

$$M(s) \leqq [1 - \rho^{-\frac{1}{2}} \int_0^1 |\, q(\xi)\,|\, d\xi]^{-1} < 2 \qquad \text{if } \rho \geqq \rho_0.$$

Similarly for u_2, using (12), so for some $\rho_0 > 0$,

(13) $|\, u_1(x,s)\sqrt{s}\, e^{-x\sqrt{s}}\,| < 2$, $|\, u_2(x,s)\sqrt{s}\, e^{-(1-x)\sqrt{s}}\,| < 2$ if $|\, s\,| = \rho \geqq \rho_0$,

and consequently these moduli are uniformly bounded in the entire s-plane.

It follows at once from (11) that

(14) $2D(s)\sqrt{s}\, e^{-\sqrt{s}} = 2u_1(1,s)\sqrt{s}\, e^{-\sqrt{s}} = 1 - e^{-2\sqrt{s}} + O(\rho^{-\frac{1}{2}})$ $(\rho \geqq \rho_0)$,

and therefore, for some finite $\gamma > 0$, none of the zeros of $D(s)$ lie in the half-plane $\mathcal{R}(s) \geqq \gamma$, because $|\, 2D(s)\sqrt{s}\, e^{-\sqrt{s}}\,|$ has a positive lower bound there.

With the aid of the above order properties it is easily seen from (11) and (12) that, in the half-plane $\mathcal{R}(s) \geqq \gamma$,

$$G_\xi(x,\xi,s) = O(|\, e^{\pm(x-\xi)\sqrt{s}}\,|) \qquad (+ \text{ if } x < \xi, - \text{ if } x > \xi),$$

and, if $\phi(\xi)$ is any bounded integrable function,

$$\int_0^1 G(x,\xi,s)\phi(\xi)\, d\xi = O(\rho^{-1}), \qquad \int_0^1 G_\xi(x,\xi,s)\phi(\xi)\, d\xi = O(\rho^{-\frac{1}{2}}).$$

Let x_i $(i = 1, 2, \cdots, m;\ 0 < x_1 < \cdots < x_m < 1)$ be the points of discontinuity of $F(x)$, and let

$$b_i = F(x_i + 0) - F(x_i - 0).$$

By setting

$$G(x,\xi,s) = s^{-1} G_{\xi\xi}(x,\xi,s) - s^{-1} q(\xi) G(x,\xi,s)$$

in (7) and integrating the first integral so obtained by parts, noting that G_ξ is discontinuous at $\xi = x$ with a jump of -1, we obtain

(15) $su(x,s) = F(x) + G_\xi(x,1,s)F(1) - G_\xi(x,0,s)F(0) - \sum_1^m b_i G_\xi(x,x_i,s)$

$$- \int_0^1 G_\xi(x,\xi,s)F'(\xi)\, d\xi - \int_0^1 G(x,\xi,s)q(\xi)F(\xi)\, d\xi,$$

if $x \neq x_i$.

To examine the behavior of u in the closed subintervals, in $0 \leqq x \leqq x_1$, for instance, we need a more precise order property of G_ξ at the endpoints. From (12), (13) and (14) we find that, for $\mathcal{R}(s) \geqq \gamma$,

$$G_\xi(x, 0, s) = -e^{-x\sqrt{s}} \frac{e^{-(1-x)\sqrt{s}} u_2(x, s)}{e^{-\sqrt{s}} D(s)} = e^{-x\sqrt{s}} \frac{1}{1 + O(\rho^{-\frac{1}{2}})} + O(\rho^{-\frac{1}{2}})$$
$$= e^{-x\sqrt{s}} + O(\rho^{-\frac{1}{2}}).$$

A similar computation gives, for $0 \leqq x \leqq x_1$ $(0 < x_1 < 1)$,

$$G_\xi(x, x_1, s) = -\tfrac{1}{2} e^{-(x_1 - x)\sqrt{s}} + O(\rho^{-\frac{1}{2}}) \qquad (\mathcal{R}(s) \geqq \gamma).$$

Recalling that u is continuous in x, equation (15) now gives the expression

$$u(x, s) = \frac{F(x) - F(0)}{s} e^{-x\sqrt{s}} + \frac{b_1}{2s} e^{-(x_1 - x)\sqrt{s}} + O(\rho^{-3/2}) \tag{16}$$

when $0 \leqq x \leqq x_1$ and $\mathcal{R}(s) \geqq \gamma$.

For every $x \geqq 0$ and every $t > 0$ it is known [7] that the inverse transform of $s^{-1} e^{-x\sqrt{s}}$ is given by

$$L^{-1}_{(i)} \left\{ \frac{e^{-x\sqrt{s}}}{s} \right\} = \frac{2}{\sqrt{\pi}} \int_{\frac{x}{2\sqrt{t}}}^{\infty} e^{-\eta^2} d\eta = I\left(\frac{x}{2\sqrt{t}}\right). \tag{17}$$

The integral I is clearly continuous in (x, t) when $t > 0$ and uniformly bounded for all $t \geqq 0$ and $x \geqq 0$. According to (16),

$$L^{-1}_{(i)}\{u(x, s)\} = -F(0) I\left(\frac{x}{2\sqrt{t}}\right) + \frac{b_1}{2} I\left(\frac{x_1 - x}{2\sqrt{t}}\right) + L^{-1}_{(i)} \left\{ \frac{F(x)}{s} + O(\rho^{-3/2}) \right\} \tag{18}$$

for $t > 0$. The last inversion integral converges, according to Lemma 1 $(n = 1)$, and approaches $F(x)$ as t approaches zero. From the character of the integral I it follows from (18) that, when $0 < x \leqq x_1$,

$$\lim_{t \to 0} L^{-1}_{(i)}\{u\} = \tfrac{1}{2}[F(x - 0) + F(x + 0)]. \tag{19}$$

The inversion integral of the function of order $\rho^{-3/2}$ converges uniformly with respect to (x, t) in each finite interval in $t \geqq 0$; it is continuous in (x, t) for each $t \geqq 0$ and $0 \leqq x \leqq x_1$ and of order $e^{\gamma t}$ for all $t \geqq 0$. Consequently $L^{-1}_{(i)}\{u\}$ has this continuity for each $t > 0$, and the same order.

When x is in either of the closed intervals $(x_1, x_2), \cdots, (x_m, 1)$ it can

[7] Part II of the paper cited under 3, *Mathematische Annalen*, vol. 115 (1938), pp. 720-739; p. 727.

be seen in the same way that $L^{-1}_{(i)}\{u\}$ has the corresponding properties. Its continuity in x at $x = 0$ and $x = 1$ $(t > 0)$ shows that it vanishes there since $u(0, s) = u(1, s) = 0$. Its continuity in (x, t) at $x = x_1, x_2, \cdots, x_m$, for $t > 0$ can easily be seen from its continuity in the closed subintervals and the continuity of $u(x, s)$; it will be shown independently in the next section. Our chief results can be stated as follows.

LEMMA 2. *Let $q(x)$ be continuous, $F(x)$ piecewise continuous and $F'(x)$ bounded and integrable, and let $u(x, s)$ be defined by equation (7). Then $L^{-1}_{(i)}\{u\}$ converges to a continuous function of (x, t) for each $t > 0$, $0 \leqq x \leqq 1$ having the transform $u(x, s)$ for $\Re(s) > \gamma$; it is uniformly of order $O(e^{\gamma t})$ when $0 \leqq x \leqq 1$, $t \geqq 0$, and satisfies the boundary conditions (3), and (19) for $0 < x < 1$, and hence (2). γ is any positive number such that the half-plane $\Re(s) \geqq \gamma$ contains no zeros of $D(s)$.*

The analytic function u is real when s is real so $u(x, \gamma - i\eta)$ is the complex conjugate of $u(x, \gamma + i\eta)$, according to the reflection principle of Schwarz. Consequently the inversion integral can be written as a real integral:[8]

$$L^{-1}_{(i)}\{u\} = \frac{e^{t\gamma}}{\pi}\int_0^\infty \Re\left(e^{it\eta}u(x, \gamma + i\eta)\right)d\eta. \tag{20}$$

By imposing additional conditions on the derivatives of F and q it is possible to continue our study of $u(x, s)$ so as to show that the inversion integral satisfies the heat equation (1), (compare Sec. 6). But we can show this without additional conditions by writing the inversion integral as a series.

4. A solution of the temperature problem, in two forms. For every fixed s the solutions u_1, u_2 of (10) satisfy the condition

$$u_1u'_2 - u_2u'_1 = \text{constant} = D(s).$$

So if $s = s_n$, a zero of $D(s)$, u_1 differs from u_2 by a constant factor; it is the characteristic function of the Sturm-Liouville system ((10), (6)) corresponding to the characteristic number s_n. Let $\phi_n(x)$ be the normalized function. The proofs that the numbers s_n are all real, and that $G(x, \xi, s)$ has simple poles at $s = s_n$ with residues $\phi_n(x)\phi_n(\xi)$, are simple and well-known.[9] According to (7), s_n is also a simple pole of $u(x, s)$; the residue there is

[8] Theorem 2 of the paper cited under 3.

[9] See, for instance, Riemann-Weber, *Differentialgleichung der Physik*, vol. I (1925), Chap. 7. The absence of the usual negative sign from the above expression for the residues of G is due to our use of the parameter s in (10) in place of the usual $-\lambda$.

$$A_n = \phi_n(x) \int_0^1 \phi_n(\xi) F(\xi)\, d\xi = c_n \phi_n(x)$$

where c_n is the Fourier constant of $F(x)$.

When $s = -\zeta^2$, ζ real, it is clear from (14) that the characteristic equation $D(s) = 0$ has the form $\sin \zeta + O(\zeta^{-1}) = 0$, so the characteristic numbers approach $-n^2\pi^2$ for large n. It is also true that it is the n-th characteristic number which has that asymptotic value,[10] so just one of the numbers lies between $-(n-\frac{1}{2})^2\pi^2$ and $-(n+\frac{1}{2})^2\pi^2$, for sufficiently large n.

Let us now determine the order of $u(x, s)$ when the points $s = \rho e^{i\theta}$ lie on the arcs P_n, in $\mathcal{R}(s) \leqq \gamma$, of the parabolas

$$\rho = (n+\tfrac{1}{2})^2\pi^2 \csc^2 \theta/2 \qquad (n = 1, 2, \cdots).$$

Just one pole of u lies between P_n and P_{n+1}, for sufficiently large n. It is easily seen from (13) that, for some $\rho_0 > 0$,

$$\sqrt{s}\, G(x, \xi, s) = O(|\sqrt{s}\, D(s) e^{-\sqrt{s}}|^{-1}) \qquad (\rho \geqq \rho_0).$$

Since $\sqrt{s} = (n+\frac{1}{2})\pi(|\cot \theta/2| \pm i)$ when s is on P_n, it follows from (14) that $|\sqrt{s}\, D(s) e^{-\sqrt{s}}| > \frac{1}{4}$ for sufficiently large n. So for some fixed N, $|\sqrt{s}\, G(x, \xi, s)| < M$ for all $n \geqq N$, where m is independent of x, ξ, and s. Consequently when s is on P_n,

$$|\sqrt{s}\, u(x, s)| < M \int_0^1 |F(\xi)|\, d\xi \qquad (n \geqq N).$$

Under the conditions just established the inversion integral of u can be expanded, according to a known theorem,[11] in the series $\sum_1^\infty A_n e^{s_n t}$, for $t > 0$, provided the integral converges. So for $0 \leqq x \leqq 1$,

$$L^{-1}_{(t)}\{u\} = \sum_1^\infty c_n \phi_n(x) e^{s_n t} \qquad (t > 0), \tag{21}$$

where

$$c_n = \int_0^1 \phi_n(\xi) F(\xi)\, d\xi.$$

We shall now show that the series in (21) satisfies the heat equation (1). It follows from (11) and (13) that $|\sqrt{s_n}\, u_1(x, s_n)|$ and $|u'_1(x, s_n)|$ are bounded uniformly for all n and x. A direct estimation from these same equations shows that the normalizing factor has the property

---

[10] A. Kneser, "Untersuchungen über die Darstellung willkürlicher Funktionen in der mathematischen Physik," *Mathematische Annalen*, vol. 58 (1904), pp. 81-147, p. 138.

[11] R. V. Churchill, "Additional notes on the inversion of the Laplace transformation," *Mathematische Zeitschrift*, vol. 43 (1938), pp. 743-748; Theorem 1.

$$\lim_{n\to\infty} |\, s_n \,| \int_0^1 u_1^2(\xi, s_n)\, d\xi = \tfrac{1}{2};$$

consequently $|\, \phi_n(x)\,|$ and $|\, s_n^{-\frac{1}{2}}\phi'_n(x)\,|$ are bounded uniformly in n and x, for all n greater than some fixed number.

For each $T > 0$ the terms of the series (21) are numerically less than $Me^{s_n T}$ for all $t \geqq T$ and $n \geqq N$, where M is a fixed positive constant and N is an integer such that $s_N < 0$. The series therefore converges absolutely and uniformly in (x, t) for $t \geqq T$, $0 \leqq x \leqq 1$, to a continuous function of these variables. This is also true of the series obtained by differentiating once with respect to either x or t. Since $\phi''_n = s_n\phi_n + q\phi_n$, the same is true for the series obtained by differentiating twice with respect to x, and so the function defined by the series satisfies the heat equation for $t > 0$.

Since this function was identified with $L_{(i)}^{-1}\{u\}$ by (21), it is a solution of our problem, according to Lemma 2.

THEOREM 1. *Under the conditions of Lemma 2, the function*

$$U(x, t) = L_{(i)}^{-1}\{u\} \equiv \sum_1^\infty c_n\phi_n(x)\, e^{s_n t} \qquad (t > 0)$$

is a solution of the boundary value problem (1)–(3); *it can also be written in the real integral form* (20). *It is of order* $O(e^{\gamma t})$ *uniformly in* x $(0 \leqq x \leqq 1)$, *for all* $t \geqq 0$ *and, together with* U_x, U_{xx} *and* U_t, *it is continuous in* (x, t) *for each* $t > 0$, $0 \leqq x \leqq 1$. *Its transform is the function* $u(x, s)$.

It is of interest to remark here that the process in this section can be extended quite easily to show that the series (21) with $t = 0$ converges to $F(x)$. The conditions on $F(x)$ and $q(x)$ given in Lemma 2 are sufficient for this expansion theorem. Since the function $u - s^{-1}F$ has an inversion integral which vanishes at $t = 0$ when $x \neq 0, x_1, \cdots, x_m, 1$, it only remains to show that the integral of that function over the arc P_n approaches zero as n becomes infinite.

5. Uniqueness theorem. If there are two solutions of our problem (1)–(3), their difference $V(x, t)$ will be a solution of the homogeneous problem

$$V_{xx} - qV = V_t \qquad (t > 0;\ 0 < x < 1), \tag{22}$$

$$V(x, +0) = 0 \qquad (0 < x < 1), \tag{23}$$

$$V(+0, t) = 0, \qquad V(1 - 0, t) \qquad (t > 0). \tag{24}$$

Conditions under which the solution of this problem will vanish identically will now be established.

THEOREM 2. *Let the solution $V(x,t)$ of the homogeneous problem (22)-(24), in which $q(x)$ is continuous when $0 \leqq x \leqq 1$, be required to satisfy each of the following conditions:*

(a) *For each $t > 0$, V and V_x exist and are continuous in (x,t) when $0 \leqq x \leqq 1$, while V_t behaves in the same way except possibly at certain points $x = 0, x_1, \cdots, x_m, 1$.*

(b) *For each $t_0 > 0$, $|V|$ is uniformly bounded when $0 \leqq t \leqq t_0$ and $0 \leqq x \leqq 1$, and $|V_t|$ has the same property for all x in each closed interval containing neither $x = 0, x_1, \cdots, x_m$, nor 1; also $\int_0^{t_0} |V_x|\,dt$ converges for each x and vanishes as $t_0 \to 0$, uniformly for all x $(0 \leqq x \leqq 1)$.*

Then $V(x,t) \equiv 0$ for all $t > 0$.

For any fixed $T > 0$ we define the function W as follows:

$$W(x,t) = V(x,t) \text{ when } 0 < t \leqq T, \qquad = 0 \text{ when } t > T.$$

Then W satisfies (22)-(24), and we shall show by means of its transform that W vanishes. For every x the transform w exists, since

$$L\{W\} = w(x,s) = \int_0^T e^{-st} V(x,t)\,dt,$$

and is continuous in x $(0 \leqq x \leqq 1)$. The conditions (a) and (b) also insure the uniform convergence and continuity of $L\{W_x\}$ for all x $(0 \leqq x \leqq 1)$; also

$$L\{W_x\} = \int_0^T e^{-st} V_x(x,t)\,dt = w_x(x,s).$$

Similarly for all x except possibly $x = 0, x_1, \cdots, x_m, 1$, it is true that $L\{W_{xx}\} = w_{xx}(x,s)$, since $W_{xx} = qW + W_t$ and therefore W_{xx} satisfies the conditions imposed on W_t under (a) and (b). Finally, in view of (23),

$$L\{W_t\} = \int_0^T e^{-st} V_t(x,t)\,dt = e^{-sT} V(x,T) + sw(x,s) \qquad (x \neq 0, x_i, 1),$$

where the integration by parts is justified by the continuity of V and the boundedness and integrability of V_t, with respect to t.

To all members of the equations (22) and (24), with V replaced by W, the operator L can therefore be applied, and we find that $w(x,s)$ must be a solution of the system.

$$w_{xx} - (q + s)w = e^{-sT} V(x,T), \qquad w(+0,s) = w(1-0,s) = 0.$$

But our function w, together with its derivative w_x, is continuous in x

$(0 \leqq x \leqq 1)$, and, for each s other than a characteristic number s_n, there is just one solution of the above system having this character, namely,

$$w(x, s) = -e^{-sT} \int_0^1 G(x, \xi, s)\, V(\xi, T)\, d\xi = -e^{-sT} y(x, s),$$

where y represents the integral. This expression for w is thus established for $\Re(s) \geqq \gamma$, where γ and G have the same significance as in Section 3.

Now $V(x, T)$ and $V_x(x, T)$ are continuous in x, so according to Lemma 2, $L_{(t)}^{-1}\{y\}$ converges to a function $Y(x, t)$ having the transform $y(x, s)$ for $\Re(s) > \gamma$. A simple property of the Laplace transformation shows that the function

$$Z(x, t) = 0 \text{ when } 0 < t < T, \quad = -Y(x, t - T) \text{ when } t > T,$$

has the transform $-e^{-sT} y(x, s)$. Since Z and W have the same transform and are continuous in t when $0 < t < T$, they are identical for those values of t. But $W = V$ when $0 < t < T$, so

$$V(x, t) \equiv 0$$

for each such t. This is true for all $t > 0$ because T was an arbitrary positive number, so the theorem is proved.

It is immediately evident from the theorem that there is not more than one function $U(x, t)$ which satisfies (1), (2), and (3) and the additional conditions (a) and (b), when $q(x)$ is continuous. Moreover, this fact does not depend upon $F(x)$ or the homogeneity of equations (1) and (3).

It is important to note that the additional conditions (a) and (b) are conditions of the type to be expected from the physical viewpoint if the solution of the problem (1)–(3) is to be unique. For condition (a) would not be satisfied if instantaneous sources of heat were introduced into the bar after $t = 0$, and (b) is not satisfied if instantaneous sources are present at $t = 0$. The lack of uniqueness when such sources are present has been discussed in the literature.[12] Again we can anticipate from the physical problem that restrictions on the derivatives of the initial temperature $F(x)$ will be needed to establish a unique solution. Such conditions will be introduced in the following section.

6. The uniqueness of the solution. We shall now prove

THEOREM 3. *Let $q(x)$ be continuous, $F(x)$ and its first two derivatives piecewise continuous, and $F'''(x)$ and $q'(x)$ bounded and integrable, when $0 \leqq x \leqq 1$. Then the solution given in Theorem* 1, *of the boundary value*

[12] See the paper cited under 1 and the book cited under 5.

problem (1)–(3), *satisfies the conditions* (a) *and* (b) *of Theorem* 2, *and it is the only solution satisfying those conditions.*

The solution $U(x,t)$ described in Theorem 1 satisfies conditions (a) and the first one of conditions (b) under the original restrictions on F and q. The additional restrictions are introduced to show that U_t and U_x satisfy the requirements in (b).

Let x_i $(i=1,2,\cdots,m)$ now represent those points at which either F, F', or F'' have discontinuities. Also let x lie in any closed interval containing neither $0, x_1, \cdots, x_m$, nor 1. By integration by parts,

$$\int_0^1 G_\xi(x,\xi,s)F'(\xi)\,d\xi = -\int_0^1 G(x,\xi,s)F''(\xi)\,d\xi + O(\rho^{-k})$$

$(\mathcal{R}(s) \geqq \gamma)$, for every real k. Hence for $\mathcal{R}(s) \geqq \gamma$, (15) can be written

$$su(x,s) = F(x) + \int_0^1 G(x,\xi,s)[F''(\xi) - q(\xi)F(\xi)]\,d\xi + O(\rho^{-k}).$$

But the integral here has the same character as $u(x,s)$ and, according to (15) itself, it equals $s^{-1}[F''(x) - q(x)F(x)] + O(\rho^{-3/2})$, so

$$u(x,s) = \frac{F(x)}{s} + \frac{F''(x) - q(x)F(x)}{s^2} + O(\rho^{-5/2}) \qquad (x \neq 0,\ x_i,\ 1).$$

Since $L_{(i)}^{-1}\{u\} = U$, Lemma 1 (with $n=2$) applies here to give

$$U_t(x,t) = L_{(i)}^{-1}\{su - F\} = F''(x) - q(x)F(x) + L_{(i)}^{-1}\{\psi\} \qquad (x \neq 0,\ x_i,\ 1), \tag{25}$$

where $\psi(x,s) = O(\rho^{-3/2})$, so the last inversion integral is $O(e^{\gamma t})$ for all $t \geqq 0$. It follows that U_t satisfies the conditions in (b): $|U_t|$ is bounded when $0 \leqq t \leqq t_0$ and x is in any interval not containing 0, x_i, 1.

By differentiating (15) the integral of $(G_{x\xi}F')$ appears; when this is integrated by parts we find

$$su_x(x,s) = F'(x) - F(0)G_{x\xi}(x,0,s) + F(1)G_{x\xi}(x,1,s) - \sum_1^m b_i G_{x\xi}(x,x_i,s) + \sum_1^m b'_i G_x(x,x_i,s) + O(\rho^{-1/2}) \tag{26}$$

when $\mathcal{R}(s) \geqq \gamma$, where x_i are now the points where either F or F' is discontinuous, and

$$b'_i = \tfrac{1}{2}[F'(x_i+0) - F'(x_i-0)].$$

To obtain the inversion integral of u_x when $0 \leqq x \leqq 1$ we must determine an asymptotic estimate of higher order for $G_{x\xi}$ in $\mathcal{R}(s) \geqq \gamma$.

From (13) and (11) we find that

$$u_1(x,s) = s^{-1/2}shx\sqrt{s} + s^{-1}\omega(x,s)\,e^{x\sqrt{s}}$$

where $|\omega(x,s)|$ is uniformly bounded in $\mathcal{R}(s) \geqq \gamma$. By substituting this in the integral in (11) a direct calculation gives an asymptotic expression for $D(s) = u_1(1,s)$:

$$2\sqrt{s}\,D(s)e^{-\sqrt{s}} = 1 + \frac{Q(1)}{2\sqrt{s}} + O(\rho^{-1}) \qquad (\mathcal{R}(s) \geqq \gamma),$$

where

$$Q(x) = \int_0^x q(\xi)\,d\xi.$$

In like manner we find the expressions, in $\mathcal{R}(s) \geqq \gamma$,

$$2u'_1(x,s)e^{-x\sqrt{s}} = 1 + e^{-2x\sqrt{s}} + \frac{Q(x)}{2\sqrt{s}}(1 - e^{-2x\sqrt{s}}) + O(\rho^{-1}),$$

$$2u'_2(x,s)e^{-(1-x)\sqrt{s}} = 1 + e^{-2(1-x)\sqrt{s}} + \frac{Q_1(x)}{2\sqrt{s}}(1 - e^{-2(1-x)\sqrt{s}}) + O(\rho^{-1}),$$

where $Q_1(x) = Q(1) - Q(x)$. Suitable expressions for $G_{x\xi}(x,\xi,s)$ can be calculated from these; we find, in particular,

$$G_{x\xi}(x,0,s) = [\tfrac{1}{2}Q(x) - \sqrt{s}\,]e^{-x\sqrt{s}} + O(\rho^{-1/2}) \qquad (0 \leqq x \leqq 1),$$

$$G_{x\xi}(x,x_1,s) = [\tfrac{1}{4}Q(x,x_1) - \tfrac{1}{2}\sqrt{s}\,]e^{-(x_1-x)\sqrt{s}} + O(\rho^{-1/2}) \qquad (0 \leqq x \leqq x_1),$$

$$G_x(x,x_1,s) = \tfrac{1}{2}e^{-(x_1-x)\sqrt{s}} + O(\rho^{-1/2}) \qquad (0 \leqq x \leqq x_1),$$

where $Q(x,x_1) = Q(x_1) - Q(x)$, and $\mathcal{R}(s) \geqq \gamma$.

We substitute these in (26) with $0 < x < x_1$, but because $u_x(x,s)$ is continuous in x the form obtained is valid for $0 \leqq x \leqq x_1$. Since [13]

$$L^{-1}_{(t)}\{s^{-1/2}e^{-x\sqrt{s}}\} = (\pi t)^{-1/2}e^{-x^2/4t} \qquad (x \geqq 0,\ t > 0),$$

the resulting expression for u_x consists of terms of order $\rho^{-3/2}$ and terms whose inversion integrals are known. Applying the inversion integral and noting its uniform convergence with respect to x, we find that

$$(27)\quad U_x(x,t) = F'(x) + \frac{F(0)}{\sqrt{\pi t}}e^{-x^2/4t} + \frac{b_1}{2\sqrt{\pi t}}e^{-(x_1-x)^2/4t} - F(0)Q(x)I\left(\frac{x}{2\sqrt{t}}\right) + H(x,x_1)I\left(\frac{x_1-x}{2\sqrt{t}}\right) + T(x,t)$$

when $0 \leqq x \leqq x_1$, $0 < x_1 < 1$, $t > 0$; $T(x,t) = L^{-1}_{(t)}\{O(\rho^{-3/2})\}$, and

$$H(x,x_1) = \tfrac{1}{2}b'_1 - \tfrac{1}{4}b_1Q(x,x_1)$$

while the integral I is defined in (17).

[13] See Appendix 2 and Theorem 2, p. 105, of the book cited under 5.

Corresponding expressions for U_x hold when x is in the closed intervals $(x_1, x_2), \cdots, (x_m, 1)$, so for each $t_0 > 0$, we see that $| U_x(x,t) | < A + Bt^{-1/2}$ when $0 < t \leqq t_0$ where A and B are positive constants. The condition that $\int_0^{t_0} | U_x | \, dt$ converges for each x and vanishes as t_0 approaches zero, uniformly in x $(0 \leqq x \leqq 1)$, is therefore satisfied and the theorem is proved.

7. Properties of the temperature function. The behavior of the temperature $U(x,t)$ at an endpoint and at a point of discontinuity of the initial temperature $F(x)$, when t is small, follows from (18). We can state it as follows:

$$\text{(A)} \quad U(x,t) = F(x) - F(0)I\left(\frac{x}{2\sqrt{t}}\right) + \frac{F(x_1+0) - F(x_1-0)}{2} I\left(\frac{x_1 - x}{2\sqrt{t}}\right) + \Omega(x,t)$$

when $0 \leqq x \leqq x_1$, where $\Omega(x,+0) = 0$ uniformly with respect to x.

According to (25), the initial rate of change of the temperature is

$$\text{(B)} \qquad U_t(x,+0) = F''(x) - q(x)F(x),$$

when x is neither 0, 1 nor a point of discontinuity of F, F' or F''.

For each x which is not an endpoint or a point of discontinuity of F or F', it follows from (27) that

$$\text{(C)} \qquad U_x(x,+0) = F'(x).$$

But (27) also gives the behavior of U_x when t is small at the more interesting points, the endpoints and points of discontinuity of F or F':

THEOREM 4. *The flux of heat at an endpoint at which F is not zero, or at a point where $F(x)$ is discontinuous, becomes infinite of order $t^{-1/2}$ as t approaches zero. More specifically,*

$$\text{(D)} \quad \lim_{t \to 0} \left[U_x(0,t) - F'(0) - \frac{F(0)}{\sqrt{\pi t}} \right] = 0,$$

$$\text{(E)} \quad \lim_{t \to 0} \left[U_x(x_i,t) - \frac{F'(x_i+0) + F'(x_i-0)}{2} - \frac{F(x_i+0) - F(x_i-0)}{2\sqrt{\pi t}} \right] = 0,$$

where x_i is a point of discontinuity of $F(x)$ or $F'(x)$.

THE UNIVERSITY OF MICHIGAN.

THE NUMBER OF CIRCLES COVERING A SET.*

By Richard Kershner.

It is well known that the "best" way to cover a given area with circles of a given radius ϵ, or to pack such circles within a given region is to place the centers of the circles on an equilateral triangle network, i. e., to circumscribe (inscribe) the circles about the hexagons of a regular hexagon network or honeycomb. This, of course, is not a precise statement, and, in fact, it is difficult to make a precise statement in this direction that is true. Roughly, the statement becomes more true as ϵ is taken smaller in relation to the area of the given region.

The most usual [1] precise statement of this fact is that the densest plane Punktgitter is that of the equilateral triangle. This statement avoids the difficulties caused by the boundedness of the bounded region but is less general than might be desired in that permissible packings or coverings are limited to those in which the centers of the circles form a Punktgitter.

The object of the present paper is to give a new and elementary proof of a precise statement in this direction; a statement involving no restriction on the nature of permissible coverings or on the nature of the given region. Specifically the statement to be proved is the following:

Theorem. *Let M denote a bounded plane point set and let $N(\epsilon)$ be the minimum number of circles of radius ϵ which can cover M. Then*

$$(\alpha) \qquad \lim_{\epsilon \to 0} \pi\epsilon^2 N(\epsilon) = (2\pi\sqrt{3}/9) \text{ meas } \bar{M}$$

where $\bar{M}$ denotes the closure of M.

Since the left side of (α) is simply the total area of the circles covering M, the constant $(2\pi\sqrt{3}/9) = 1.209\cdots$ may be thought of as measuring the proportion of unavoidable overlapping.

In the proof of the theorem there will be needed a number of lemmas, several of which have an independent interest. For instance, Lemma 5 and Lemma 6 together constitute, in one sense, a considerable refinement of (α) as providing an estimate for the error term implied by the limiting relation

* Received December 13, 1938.

[1] Cf., e. g., Hilbert, Cohn-Vossen, *Anschauliche Geometrie*, p. 32.

(α); but, of course, this refinement is obtained only by the sacrifice of the generality of the region M.

A proof of the known [2] Lemma 1 has been included for the sake of completeness.

LEMMA 1. *Let Γ denote a bounded plane network [3] consisting of a finite number of finite polygons. Suppose that each vertex of Γ is on at least three edges. Then the average number of sides of the polygons of Γ is < 6.*

Proof. Let V, E, F denote the number of vertices, edges and faces respectively, in Γ. Then the Euler relation for Γ is

$$V - E + F = 1. \tag{1}$$

The fact that each vertex is on at least three edges may be expressed, in view of the fact that each edge is on exactly two vertices, by the relation

$$3V \leqq 2E. \tag{2}$$

Elimination of V between (1) and (2) gives

$$E \leqq 3F - 3. \tag{3}$$

Now let the faces of Γ be numbered and let e_i $(i = 1, 2, \cdots, F)$ denote the number of edges on the face numbered i. Then, since some, but not all, edges are on two faces,

$$\sum_{i=1}^{F} e_i < 2E \leqq 6F - 6.$$

Thus,

$$(1/F) \sum_{i=1}^{F} e_i < 6 - 6/F < 6.$$

This completes the proof of Lemma 1.

The next two lemmas that are needed are of an extremely elementary nature

LEMMA 2. *Let σ be a fixed circle and let A_k denote the area of a regular polygon of k sides inscribed in σ. Then*

$$0 < A_{k+1} - A_k < A_k - A_{k-1} \qquad (k \geqq 3).$$

Proof. The first of these inequalities is well known. To show that the second inequality of Lemma 2 holds, let

---

[2] Cf., e. g., M. Goldberg, "The isoperimetric problem for polyhedra," *Tohoku Mathematical Journal*, vol. 40 (1934), pp. 228-229.

[3] By a network is meant the figure consisting of a set of non-overlapping polygons which together cover a simply connected domain.

$$f(k) = A_{k+1} - A_k = \frac{k+1}{2} r^2 \sin \frac{2\pi}{k+1} - \frac{k}{2} r^2 \sin \frac{2\pi}{k}.$$

Then

$$\frac{df(k)}{dk} = \frac{r^2}{2}\left[\sin \frac{2\pi}{k+1} - \frac{2\pi}{k+1} \cos \frac{2\pi}{k+1}\right] - \frac{r^2}{2}\left[\sin \frac{2\pi}{k} - \frac{2\pi}{k} \cos \frac{2\pi}{k}\right].$$

But this expression is negative for $k \geqq 3$ since $\sin x - x \cos x$ is an increasing function of x in $(0, \pi)$. Thus $f(k)$ is a decreasing function of k for $k \geqq 3$ which completes the proof.

LEMMA 3. *With the notations of Lemma 2 and* $k > j \geqq 3$,

$$(k-j)(A_k - A_{k-1}) \leqq A_k - A_j \leqq (k-j)(A_{j+1} - A_j).$$

Proof. This is an immediate corollary to Lemma 2 since Lemma 2 states that the interval from A_j to A_k consists of $(k-j)$ subintervals of which the longest is $A_{j+1} - A_j$ and the shortest $A_k - A_{k-1}$.

LEMMA 4. *Let* Γ *denote a bounded plane network consisting of* F *finite polygons. Suppose that each vertex of* Γ *is on at least three edges. Suppose, finally, that each polygon of* Γ *can be covered by a circle* σ *of fixed radius* r. *Then the total area of* Γ *is* $< FA_6$, *where* $A_6 = (3\sqrt{3}/2)r^2$ *is the area of a regular hexagon inscribed in* σ.

Proof. Let n_i $(i = 3, 4, \cdots, m)$ denote the number of polygons in Γ with i edges, so that $n_3 + n_4 + \cdots + n_m = F$. Then, according to Lemma 1, the average number of sides of the polygons of Γ is < 6, i. e.,

$$\sum_{i=3}^{m} i n_i < 6 \sum_{i=3}^{m} n_i$$

which may also be written

$$\sum_{i=7}^{m} (i-6) n_i < \sum_{i=3}^{5} (6-i) n_i. \tag{4}$$

By Lemma 2, if the left side of (4) be multiplied by $A_7 - A_6$ and the right by $A_6 - A_5$ the inequality will be strengthened. Thus

$$\sum_{i=7}^{m} (i-6)(A_7 - A_6) n_i < \sum_{i=3}^{5} (6-i)(A_6 - A_5) n_i. \tag{5}$$

Now, by the first inequality of Lemma 3,

$$(6-i)(A_6 - A_5) \leqq A_6 - A_i \qquad (i < 6); \tag{6}$$

and by the second,

$$A_i - A_6 \leqq (i-6)(A_7 - A_6) \qquad (i > 6). \tag{7}$$

Then from (5), (6) and (7),

$$\sum_{i=7}^{m} (A_i - A_6) n_i < \sum_{i=3}^{5} (A_6 - A_i) n_i$$

or

$$\sum_{i=3}^{m} A_i n_i < A_6 \sum_{i=3}^{m} n_i = F A_6. \tag{8}$$

Since the left member of the inequality (8) is $\geqq$ the total area of Γ the proof of Lemma 4 is complete.

LEMMA 5. *Let ρ denote a rectangle in the plane with area R. Let $N(\epsilon)$ denote the minimum number of circles of radius ϵ which can cover ρ. Then*

$$\pi\epsilon^2 N(\epsilon) > (2\pi\sqrt{3}/9)(R - 2\pi\epsilon^2).$$

Proof. Let σ_i $(i = 1, 2, \cdots, N(\epsilon))$ be $N(\epsilon)$ circles of radius ϵ which cover ρ and let C_i denote the center of σ_i. If the circles σ_i and σ_j intersect let c_{ij} denote their common chord. If the circles σ_i and σ_j are disjoint or tangent it will be said that c_{ij} does not exist.

Now, for every $i = 1, \cdots, N(\epsilon)$, let π_i be defined as the set of points Q in $\rho \cdot \sigma_i$ satisfying

$$d(Q, C_i) \leqq d(Q, C_j) \qquad \text{for any } j \neq i; \tag{9}$$

where $d(O, Q)$ is the distance between the points O and Q. The set π_i defined by (9) is clearly a closed subset of σ_i for every i. Furthermore, π_i is non-empty for every i. This is clear if C_i is in ρ, but in any case there must exist some point Q in $\rho \cdot \sigma_i$ which is in no σ_j with $j \neq i$; otherwise the $N(\epsilon) - 1$ circles σ_j, $j \neq i$, would cover ρ, contradicting the assumption that $N(\epsilon)$ was the minimum number of circles which could cover ρ.

Let B denote a boundary point of π_i. Then either B is a boundary point of ρ or

$$d(B, C_i) = d(B, C_j) \qquad \text{for some } j \neq i.$$

Thus B is either on the boundary of ρ or on some c_{ij}, so that π_i is a polygon bounded by subsegments of the c_{ij} and subsegments of the boundary of ρ.

The set of polygons π_i form a non-overlapping collection of polygons which exactly cover ρ. Unfortunately this collection of polygons does not quite satisfy the conditions of Lemma 4 since the four vertices of ρ are vertices of some π_i where only two edges meet. If these four vertices are "straightened out" a new collection of polygons π'_i results which "almost" covers ρ and, in particular, having a total area $> R - 2\pi\epsilon^2$. It will now be shown that the network consisting of the π'_i satisfies the conditions of Lemma

4 with $r = \epsilon$. To do this it is clearly sufficient to show that the network consisting of the π_i satisfy these conditions save for the four exceptional vertices mentioned. Since the second hypothesis of Lemma 4 is obviously satisfied, all that is required is to show that each vertex of any π_i, not a vertex of ρ, is on at least three edges. To this end let V be a vertex of π_k and let it be supposed first that the two edges of π_k intersecting at V are both subsegments of chords c_{ki} and c_{kj}. Then, by the definition of the c_{kh} together with the definition (9) of π_k,

$$d(V, C_k) = d(V, C_i) = d(V, C_j) \leqq d(V, C_h) \quad \text{for any } h = 1, 2, \cdots, N(\epsilon).$$

But these same relations imply that V is a point of π_i and a point of π_j so that at least three polygons, and, consequently, at least three edges, meet at V. The other case to be considered is the case that one of the edges of π_k through V is a segment of the boundary of ρ, the other edge being, of course, a subsegment of some c_{ki}. In this case, only two polygons can be shown by the above reasoning to meet at V, namely, π_k and π_i. But since the exterior of ρ also has V as a limit point there are still at least three regions about V and three edges (two of which will be segments of the boundary of ρ) meeting at V. Thus the assumptions of Lemma 4 are satisfied by the network of the π'_i.

Applying Lemma 4 it is seen that

$$R - 2\pi\epsilon^2 < N(\epsilon)\,(3\sqrt{3}/2)\,\epsilon^2$$

or

$$\pi\epsilon^2 N(\epsilon) > (2\pi\sqrt{3}/9)\,(R - 2\pi\epsilon^2).$$

This completes the proof of Lemma 5.

LEMMA 6. *Let ρ denote a rectangle in the plane with area R and perimeter p. Let $N(\epsilon)$ denote the minimum number of circles of radius ϵ which cover ρ. Then*

$$\pi\epsilon^2 N(\epsilon) < (2\pi\sqrt{3}/9)\,(R + 2p\epsilon + 16\epsilon^2).$$

Proof. Let $\tau(\epsilon)$ denote the regular hexagon inscribable in a circle of radius ϵ. Let the plane be "paved" with regular hexagons $\tau_i(\epsilon)$ and let $\tau_1(\epsilon), \tau_2(\epsilon), \cdots, \tau_N(\epsilon)$ be the hexagons of this "paving" which have a point in common with ρ. Then the set of hexagons $\tau_i(\epsilon)$ $(i = 1, 2, \cdots, N)$, cover ρ and are covered by the rectangle obtained by adjoining a linear strip of width 2ϵ all around ρ. This latter rectangle has area $R + 2p\epsilon + 16\epsilon^2$, so that

$$(3\sqrt{3}/2)\,\epsilon^2 N < R + 2p\epsilon + 16\epsilon^2. \tag{10}$$

But since $\tau_1(\epsilon) + \tau_2(\epsilon) + \cdots + \tau_N(\epsilon)$ covers ρ, the N circles of radius ϵ which circumscribe these $\tau_i(\epsilon)$ cover ρ, i. e.,

$$N(\epsilon) \leqq N. \tag{11}$$

Combining (10) and (11) completes the proof of Lemma 6.

Proof of the theorem. Let M be an arbitrary bounded plane point set and let $N(\epsilon)$ denote the minimum number of circles of radius ϵ which can cover M. Let $\bar{M}$ denote the closure of M.

It is well known that since $\bar{M}$ is a closed, bounded, plane point set, there is for every $\eta > 0$, a finite number $K = K(\eta)$ of rectangles ρ_i of area R_i and perimeter p_i $(i = 1, 2, \cdots, K)$ with the following properties:

$$\rho_1 + \rho_2 + \cdots + \rho_k \supset \bar{M} \tag{12}$$

$$\text{meas } \bar{M} \leqq R_1 + R_2 + \cdots + R_K \leqq \text{meas } \bar{M} + \eta. \tag{13}$$

Let $N_i(\epsilon)$ $(i = 1, 2, \cdots, K)$ be the minimum number of circles of radius ϵ covering ρ_i. Then, by (12),

$$N(\epsilon) \leqq N_1(\epsilon) + N_2(\epsilon) + \cdots + N_K(\epsilon).$$

Thus, by Lemma 6,

$$\pi\epsilon^2 N(\epsilon) < (2\pi\sqrt{3}/9)\left[\sum_{i=1}^{k} R_i + 2\epsilon \sum_{i=1}^{k} p_i + 16K\epsilon^2\right]$$

or, by (13),

$$\pi\epsilon^2 N(\epsilon) < (2\pi\sqrt{3}/9)\left[\text{meas } \bar{M} + \eta + 2\epsilon P(\eta) + 16\epsilon^2 K(\eta)\right]$$

where $P(\eta) = p_1 + p_2 + \cdots + p_K$. Thus

$$\limsup_{\epsilon \to 0} \pi\epsilon^2 N(\epsilon) \leqq (2\pi\sqrt{3}/9)\,(\text{meas } \bar{M} + \eta).$$

But since $\eta > 0$ was arbitrary, this is equivalent to

$$\limsup_{\epsilon \to 0} \pi\epsilon^2 N(\epsilon) \leqq (2\pi\sqrt{3}/9)\ \text{meas } \bar{M}. \tag{14}$$

Now let $\sigma_i(\epsilon)$ $(i = 1, 2, \cdots, N(\epsilon))$ be $N(\epsilon)$ circles of radius ϵ which cover M, and consequently $\bar{M}$. Let T be the set of all points of $\rho_1 + \rho_2 + \cdots + \rho_K$ which are not points of $\sigma_1(\epsilon) + \sigma_2(\epsilon) + \cdots + \sigma_{N(\epsilon)}(\epsilon)$. This set T, which by (13) has measure at most η, is clearly composed of a finite number of connected open regions. Thus there are a finite number $L = L(\eta)$ of rectangles ρ'_i $(i = 1, 2, \cdots, L)$ of total area 2η which cover T. Now let U be the set of points of $\rho_1 + \rho_2 + \cdots + \rho_K$ which are not points of $\rho'_1 + \rho'_2 + \cdots + \rho'_L$. This set U is again composed of a finite number $J = J(\eta)$ of non-overlapping rectangles ρ_i'' with areas R_i'' and perimeters p_i'' $(i = 1, 2, \cdots, J)$. These ρ_i'' have the following properties:

$$\rho_1'' + \rho_2'' + \cdots + \rho_J'' \subset \sigma_1(\epsilon) + \sigma_2(\epsilon) + \cdots + \sigma_{N(\epsilon)}(\epsilon) \tag{15}$$

$$R_1'' + R_2'' + \cdots + R_J'' \geqq \text{meas } \bar{M} - 2\eta. \tag{16}$$

Let ρ_i''' $(i = 1, 2, \cdots, J)$ denote the rectangle of area R_i''' obtained by removing a linear strip of width 2ϵ [4] all around ρ_i'', so that

$$R_i''' = R_i'' - 2\epsilon p_i'' + 16\epsilon^2 > R_i'' - 2\epsilon p_i'' + 2\pi\epsilon^2. \tag{17}$$

Let $N_i''(\epsilon)$ denote the number of circles $\sigma_j(\epsilon)$ $(j = 1, 2, \cdots, N(\epsilon))$ which are completely within ρ_i''. The $N_i''(\epsilon)$ circles $\sigma_j(\epsilon)$ within ρ_i'' obviously cover ρ_i''', in view of (15), so that, by Lemma 5,

$$\pi\epsilon^2 N_i''(\epsilon) > (2\pi\sqrt{3}/9)(R_i''' - 2\pi\epsilon^2)$$

or, by (17),

$$\pi\epsilon^2 N_i''(\epsilon) > (2\pi\sqrt{3}/9)(R_i'' - 2\epsilon p_i''). \tag{18}$$

But, by the definition of $N_i''(\epsilon)$,

$$N(\epsilon) \geqq N_1''(\epsilon) + N_2''(\epsilon) + \cdots + N_J''(\epsilon).$$

Thus, (18) implies

$$\pi\epsilon^2 N(\epsilon) > (2\pi\sqrt{3}/9)\left[\sum_{i=1} R_i'' - 2\epsilon \sum_{i=1}^{J} p_i''\right]$$

or, by (16),

$$\pi\epsilon^2 N(\epsilon) > (2\pi\sqrt{3}/9)[\text{meas } \bar{M} - 2\eta - 2\epsilon P''(\eta)]$$

where $P''(\eta) = p_1'' + p_2'' + \cdots + p_J''$. Thus

$$\liminf_{\epsilon \to 0} \pi\epsilon^2 N(\epsilon) \geqq (2\pi\sqrt{3}/9)(\text{meas } \bar{M} - 2\eta).$$

But, since $\eta > 0$ was arbitrary this is equivalent to

$$\liminf_{\epsilon \to 0} \pi\epsilon^2 N(\epsilon) \geqq (2\pi\sqrt{3}/9) \text{ meas } \bar{M}. \tag{19}$$

The two inequalities (14) and (19) together imply the desired equality (α).

THE UNIVERSITY OF WISCONSIN.

[4] The ρ_i'' being fixed it is supposed that ϵ is chosen to be sufficiently small for this to be possible, for all $i = 1, 2, \cdots, J = J(\eta)$.

ON BUFFON'S PROBLEM AND ITS GENERALIZATIONS.*

By M. KAC,** E. R. VAN KAMPEN and AUREL WINTNER.

1. If G denotes the grating formed by the sequence of lines

$$G: \qquad x = n;\ n = 0, \pm 1, \pm 2, \cdots \tag{1}$$

in an (x, y)-plane, what is the probable number of common points of G and of a plane convex curve C, if one throws C at random on G? A refinement of this question is as follows: What is the distribution of the number of common points of C with G, if C is thought of as situated on G in the (x, y)-plane? (The momentum of the first order (expectation) belonging to this distribution represents the answer to the first question).

The main point in these questions is, not a calculation leading to the probabilities involved, but rather the fact that the choice of a probability distribution, a choice which is necessary in order to make the above questions meaningful, may be analyzed in such a way as to clear up the tacit assumptions which underlie the usual treatment of Buffon's needle problem and its classical generalizations.[1]

2. Before throwing C on the (x, y)-plane, mark on the plane, S, of C a point P and a direction D through P.

After C has been thrown on G, the position of C in the (x, y)-plane is uniquely determined by the Cartesian coördinates, x and y, of P and by the inclination, say $2\pi\theta$ $(0 \leqq \theta < 1)$, of D. The ordinate y of P, being measured in the direction parallel to the grating G, will not explicitly enter into the following considerations; while x will only enter as reduced mod 1, i. e., as

$$\phi = \{x\}, \text{ where } \{x\} = x - [x], \qquad (0 \leqq \phi < 1), \tag{2}$$

since the distance between two subsequent lines of G is 1. Thus, the space of all possible positions of C relative to G is in one-to-one continuous correspondence with the points of the torus

$$T: \qquad 0 \leqq \phi < 1, \qquad 0 \leqq \theta < 1 \qquad \text{[cf. (2)]}; \tag{3}$$

so that the space of all such positions will be considered as identical with (3).

* Received February 1, 1939.

** Fellow of the Parnas Foundation, Lwów, Poland.

[1] Cf., e. g., J. L. Coolidge, "An introduction to mathematical probability," Oxford (1925), pp. 80-83.

At first glance, it seems to be natural to define the probability measure (with reference to which the questions raised in § 1 are to be answered) in the following manner: The probability $\mu = \mu(E)$ that a given position (ϕ, θ) of C with reference to G will lie in an open set E of T is the Euclidean (ϕ, θ)-area of E. Actually, this choice of the distribution function $\mu(E)$ of the probabilities on T cannot be accepted without some justification (cf. "Bertrand's paradox").[2]

In order to select among all probability distributions μ (which are arbitrary in view of Bertrand's "paradox") those which fit the physical experiment of throwing C on G, one will be inclined to require that

(i) the probability distribution $\mu(E)$ on T is, for every fixed P, invariant under arbitrary rotations of the fixed direction through P in the plane S;

(ii) the variables ϕ and θ are statistically independent for every choice of P in S;

(iii) the probability distribution $\mu(E)$ on T is invariant under all translations of P in the plane S of C.

Actually, it will turn out that

(I) the requirement (i) implies (ii);

(II) the requirement (iii) implies (ii); and (ii) implies (iii), unless the probability distribution of the angular variable θ is concentrated (in an arbitrary manner) on a finite number, say k, of equidistant points on the θ-circle, while the distribution of ϕ is periodic with period $1/k$;

(III) any of the requirements (i), (ii), (iii) implies that the measure on T is a product measure; and that the distribution of ϕ is the angular equidistribution, unless the requirement is (ii) and one has to do with the exceptional case noted under (II);

(IV) the requirement (i) implies, in addition, that the distribution of θ is the angular equidistribution.

Now, (I), (II), (III), (IV) clearly imply the following theorem:

If (i) is satisfied, the probability measure $\mu(E)$ on T is the classical one, i. e., the Euclidean (ϕ, θ)-measure on T.

3. It will be sufficient to prove this consequence of the four statements (I)-(IV). In fact, it will be clear from the proof how the sharper result which is contained in (I)-(IV) may be established.

---

[2] Cf. *loc. cit.*[1], pp. 75-76.

First, (i) implies that $\mu(E)$ is invariant under any rotation $\bar{\theta} = \theta +$ const. of the torus T. Hence, $\mu(E)$ must be a product measure; i. e., $\mu(E) = \mu_1(E_1)\,\mu_2(E_2)$, if the Borel set E on T is the product $E_1 \times E_2$ of Borel sets $E_1; E_2$ on the respective circles $0 \leqq \phi < 1$; $0 \leqq \theta < 1$, and $\mu_1; \mu_2$ denote certain distribution functions on these circles.[3] Since it is clear that $\mu_2(E_2)$ must also be invariant on the rotations $\bar{\theta} = \theta +$ const. of T, the distribution function $\mu_2(E_2)$ must be the angular equidistribution, i. e., $\mu_2(E_2)$ is the Lebesgue measure $| E_2 |$ of E_2. On the other hand, the fact that μ is a product measure $\mu_1\mu_2$ on T may be expressed by saying that the angular variables ϕ and θ are statistically independent. Finally, the assumption of (i), that P may be chosen arbitrarily in the plane S of C, can only be satisfied if, not only the angular variables ϕ and θ are statistically independent, but so are $\phi + l\cos(\alpha + \theta)$ and θ also, where the length l and the angle α are arbitrary. Consequently,[4] $\mu_1(E_1)$ is the Lebesgue measure $| E_1 |$ of E_1. This completes the proof of $\mu(E_1 \times E_2) = | E_1 |\,| E_2 |$.

4. The distribution function of probabilities which is meant in the questions raised in § 1 is now specified. The answer to these questions, which thus become meaningful, may now readily be obtained, as follows:

Choose the point P in S and the direction D through P as fixed. Let $f(\vartheta)$ be the supporting function of C, where $2\pi\vartheta$ $(0 \leqq \vartheta < 1)$ is the inclination of D towards an arbitrarily fixed direction in the plane S of C. Using as probability distribution on T the equidistribution $\mu(E_1 \times E_2) = | E_1 |\,| E_2 |$, one readily sees that the set of those points of T to which there correspond positions of C with reference to G such that at least one line of G is a supporting line of C, must be a set of measure zero. Since all probabilities involved are defined in terms of the measure μ, this 0-set may be disregarded. Consequently, it is sufficient to consider those positions of C with reference to G in which any line of G has on the convex curve C either no point or exactly two points. But then the number $N = N(\phi, \theta)$ of common points of C and G in that position of C which is determined by the point (ϕ, θ) of (3) is readily seen to be [5]

---

[3] The truth of the statement $\mu(E_1 \times E_2) = \mu_1(E_1)\,\mu_2(E_2)$ readily follows by choosing the ϕ-set E_1 arbitrarily, the θ-set E_2 so as to be an interval of length $1/n$, and then observing that, in view of the invariance assumption, $\mu(E_1 \times E_2)$ is independent of the particular choice of the θ-interval E_2 of length $1/n$.

[4] Cf. M. Kac and E. R. van Kampen, *American Journal of Mathematics*, vol. 61 (1939), pp. 677-682.

[5] Notice that either of the expressions [ ] occurring in (4) may be negative.

$$N(\phi,\theta)=2(1+[f(-\theta)+\phi]+[f(\tfrac{1}{2}-\theta)-\phi]), \quad (0\leqq\theta<1,\ \ 0\leqq\phi<1), \tag{4}$$

where $[g]$ denotes the largest integer not exceeding g.

It follows that the distribution function $\sigma(\xi), -\infty<\xi<+\infty$, which represents the probability that there are on C at least ξ points of G is represented by the Euclidean (ϕ,θ)-measure of the set of those points of T at which the function $N(\phi,\theta)$ does not exceed ξ (in particular, $\sigma(\xi)$ is a step function whose jumps occur at a finite number of non-negative even integer values of ξ). Consequently, the Fourier-Stieltjes transform

$$L(u)=\int_{-\infty}^{+\infty}\exp(iu\xi)\,d\sigma(\xi) \tag{5}$$

of $\sigma(\xi)$ may be written in the form

$$L(u)=\int_0^1\int_0^1\exp\,(2(1+[f(-\theta)+\phi]+[f(\tfrac{1}{2}-\theta)-\phi])ui)\,d\phi d\theta. \tag{6}$$

Clearly, the probable number of common points of C and G, i. e., the first momentum

$$M=\int_{-\infty}^{+\infty}\xi\,d\sigma(\xi) \tag{7}$$

of $\sigma(\xi)$, is

$$M=2+2\int_0^1\int_0^1([f(-\theta)+\phi]+[f(\tfrac{1}{2}-\theta)-\phi])\,d\phi d\theta. \tag{8}$$

5. As an illustration, consider the case of a circle C of given radius r. In this case, $\sigma(\xi)$ may be obtained, without recourse to (6) and the inversion formula of the transform (5), by the above description of $\sigma(\xi)$ as the Euclidean area of a portion of T. In fact, on substituting the supporting function $f(\vartheta)=r$ of C into (4), one readily finds that

$$\begin{aligned} &w(2[r])=1-2\{r\} \text{ and } w(2[r]+1)=2\{r\}, && \text{if } 0\leqq\{r\}\leqq\tfrac{1}{2};\\ &w(2[r]+1)=2-2\{r\} \text{ and } w(2[r]+2)=2\{r\}-1, && \text{if } \tfrac{1}{2}\leqq\{r\}\leqq 1, \end{aligned} \tag{9}$$

where $\{r\}=r-[r]$, and $w=w(n)$ denotes, for $n=0,1,2,\cdots$, the probability that exactly n lines of the grating (1) intersect the circle C. Needless to say, (9) may be verified directly.

As another example, consider the classical needle problem of Buffon, i. e., the case in which the convex curve consists of a double segment of given length, say $2l$. Then the supporting function of C is $f(\vartheta) = |\, l \cos 2\pi\vartheta \,|$. Substituting this into (4) and then using (6), one obtains

$$(10)\quad L(u) = \int_0^1 \int_0^1 \exp(iu(1 + [\,|\, l\cos 2\pi\theta \,| + \phi] + [\,|\, l\cos 2\pi\theta \,| - \phi]))\, d\phi d\theta,$$

if, corresponding to the fact that the curve C is a double segment, one omits the factor 2 in the representation (4) of $N(\phi, \theta)$.

6. The classical problem of Buffon, as presented in text-books, concerns, not the probable number of intersections, or the associated distribution function, for a fixed length of the needle, but rather the probability that the number of intersections of the needle and G be distinct from 0. This question may, of course, be answered by calculating $1 - \sigma(+0)$.

A corresponding consequence of the formulae of § 4 in case of a convex curve of arbitrary shape leads to well-known formulae of Cauchy,[6] as follows:

Suppose that the dimensions of the convex curve C are so small that the supporting function $f(\vartheta)$ satisfies the inequality

$$f(\vartheta) + f(\tfrac{1}{2} + \vartheta) \leqq 1$$

for every ϑ. Then the probability, $1 - \sigma(+0)$, that the number of intersections of C and G be distinct from 0 is $1 - \mu(E)$, where E is the set of those points (θ, ϕ) of T at which

$$1 + [f(-\theta) + \phi] + [f(\tfrac{1}{2} - \theta) - \phi] = 0.$$

This gives

$$1 - \sigma(+0) = 1 - \int_0^1 \Big(\int_{f(\frac{1}{2}-\theta)}^{1-f(-\theta)} d\phi \Big) d\theta = 2 \int_0^1 f(\theta)\, d\theta.$$

Since the last integral is the length of C, the result reduces to that of Cauchy.[6]

THE JOHNS HOPKINS UNIVERSITY.

---

[6] A. Cauchy, *Oeuvres*, sér. I, vol. 2, pp. 167-177; cf. H. Steinhaus, *Leipziger Sitzber.*, vol. 82 (1930), pp. 120-130.

CIRCULAR EQUIDISTRIBUTIONS AND STATISTICAL INDEPENDENCE.*

By M. KAC ** and E. R. VAN KAMPEN.

It is known that the Gaussian distribution of linear variables may be characterized by several requirements imposed on the distribution, requirements which involve at least one of the notions stability, independence, indefinite iteration, and so on.

Wintner has pointed out [1] that the angular equidistribution is a Gaussian distribution (mod 1) of vanishing precision; Lévy [2] recently remarked that the requirement of stability either leads to a circular equidistribution or to a discontinuous distribution of which the probability is equally distributed over a finite number of equidistant points of the circle, and gave several other closely related characterizations of the same distribution functions. However, in the literature of the subject there seems to be some disagreement [3] as to what should play the rôle of the Gaussian distribution in case the variables are angular instead of linear. In fact, while the different possible approaches lead in the linear case to one and the same distribution (namely the Gaussian), such is not the case for angular variables.

The object of the present paper is to show that, except in a very special case, the circular equidistribution results when the requirement of stability is replaced by an independence requirement depending on two functions.

This characterization of the equidistributions (to which we were led by the question of the independence of angle and midpoint coördinate in Buffon's needle problem [4]) depends on the answer to the following question concerning measure-preserving mappings:

If there is given on the torus Φ: $0 \leqq \phi_1 < 1$, $0 \leqq \phi_2 < 1$, a product-measure μ satisfying $\mu(\Phi) = 1$, and if T: $\bar{\phi}_1 \equiv \phi_1 + f(\phi_2) \pmod 1$, $\bar{\phi}_2 = \phi_2$ is a one-to-one transformation of Φ into itself, what are those functions f for

* Received February 1, 1939.

** Fellow of the Parnas Foundation, Lwów, Poland.

[1] A. Wintner, "On the stable distribution laws," *American Journal of Mathematics*, vol. 55 (1933), p. 339.

[2] P. Lévy, "L'addition des variables aléatoires définies modulo un," *Comptes Rendus*, vol. 207 (1938), pp. 444-446.

[3] Cf. e. g., *Handbuch der Physik*, vol. III, pp. 476-478.

[4] Cf. M. Kac, E. R. van Kampen and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 672-676.

which the mapping T preserves the measure μ? It turns out that the answer is that these functions f must have a circular equidistribution unless μ is of a very special type. This measure theoretical formulation of the question, which will not be considered in detail, does not seem to lend itself to a direct attack. However, if the problem is subjected to a suitable Fourier transformation, it reduces to a functional equation of the form

$$\int_0^1 h(t) \exp(iug(t))\, dt = 0, \qquad -\infty < u < +\infty,$$

where h and g are measurable functions, the first of which is integrable. This equation implies that the integral of $h(t)$ must vanish over the set of t-values defined by $g(t) < \omega$, where ω is any fixed real number.

A remark in **4.** concerns the characterization by means of independent functions of a different type of angular distribution.

1. A real measurable function $\phi = \phi(x)$ on the interval $0 \leqq x \leqq 1$ will be called an angular function if its values ϕ are to be reduced mod a, where a is a fixed number. It will always be assumed that $a = 1$. Thus, $\{f(x)\}$, where $\{z\} = z - [z]$, is an angular function for every real measurable $f(x)$ on $0 \leqq x \leqq 1$. Correspondingly, $\{f_1(x_1, x_2) + (f_2(x_1, x_2)\}$, where both f_1 and f_2 are measurable functions on the square $0 \leqq x_1 \leqq 1$, $0 \leqq x_2 \leqq 1$, will be called the angular sum of the two functions f_1 and f_2. The statistical independence of two functions f_1, f_2 will be meant in the sense of Steinhaus,[5] with an evident modification of his definition in case at least one of the functions is angular. The distribution function of an angular variable will be meant in a corresponding sense.

It is understood that, in case of angular variables, the Fourier-Stieltjes transform of the linear distribution function must be replaced by the sequence of the Fourier-Stieltjes coefficients of the angular distribution function. In particular, two real measurable functions $f_1(x_1, x_2)$, $f_2(x_1, x_2)$ on the square $0 \leqq x_1 \leqq 1$, $0 \leqq x_2 \leqq 1$ are statistically independent in the angular sense if and only if, on placing $e(z) = \exp(2\pi i z)$, one has

$$\int_0^1\int_0^1 e(mf_1 + nf_2)\, dx_1 dx_2 = \int_0^1\int_0^1 e(mf_1)\, dx_1 dx_2 \int_0^1\int_0^1 e(nf_2)\, dx_1 dx_2$$

[5] Cf., e. g., M. Kac, "Sur les fonctions indépendantes, I," *Studia Mathematica*, vol. 6 (1936), pp. 46-58.

for arbitrary integers m, n. A similar criterion holds if only one of the two functions f_1, f_2 is considered as angular.

2. By the angular equidistribution will be meant the angular distribution function of constant density $(=1)$.

Theorem I: For two angular functions $\phi(x)$, $\psi(y)$, where $0 \leqq x \leqq 1$, $0 \leqq y \leqq 1$, put

$$\chi(x, y) = \phi(x) + \psi(y), \text{ where } \{z\} = z - [z]. \tag{1}$$

Then $\chi(x, y)$ is statistically independent of y if and only if either

(i) the angular distribution of the function $\phi(x)$ is the angular equidistribution, while the distribution of $\psi(y)$ is arbitrary; or

(ii) for a certain integer $k \geqq 1$, the probability distribution of $\phi(x)$ over the circle $0 \leqq \phi < 1$ is periodic with primitive period $1/k$, while for a number α, $0 \leqq \alpha < 1/k$, the probability is 1 that $\psi(y)$ attains one of the values $l/k + \alpha$; $l = 0, \cdots, k$.

In order to prove this, notice first that the statistical independence of $\chi(x, y)$ and y, when expressed in terms of Fourier-Stieltjes transforms, may be written in the form

$$\int_0^1 \int_0^1 e(n\chi(x, y) + uy)\, dxdy = \int_0^1 \int_0^1 e(n\chi(x, y))\, dxdy \cdot \int_0^1 e(uy)\, dy,$$

where the integer n and the real parameter u are arbitrary. This condition is, in view of (1), equivalent to

$$A(n)\, B(n, u) = A(n)\, B(0, u)\, B(n, 0), \tag{2}$$

where $A(n)$, $B(n, u)$ are abbreviations for

$$A(n) = \int_0^1 e(n\phi(x))\, dx, \qquad B(n, u) = \int_0^1 e(n\psi(y) + uy)\, dy. \tag{3}$$

Two cases will be distinguished, according as one does or does not have

$$A(n) = 0 \text{ for every } n \neq 0. \tag{4}$$

In the first case, (2) is obvious from (3) or from (4) according as $n = 0$ or $n \neq 0$. On the other hand, the assumption (4) implies, in view

of (3) and the uniqueness theorem of the trigonometic momentum problem, that $\phi(x)$ has the distribution function stated under (i).

In the second case, where (4) does not hold, there exists at least one integer l such that

$$A(l) \neq 0, \quad l > 0. \tag{5}$$

Hence, from (2),

$$B(l, u) = B(0, u)\, B(l, 0). \tag{6}$$

This equation is satisfied if $e(\psi(y)) = \text{const.}$ for almost all y. It will be proved in § 3 that no other solution of (6) exists.

Assuming that this has already been established, one has $e(l\psi(y)) = \text{const.}$ almost everywhere for all integers l satisfying (5). If $k \geqq 1$ is the greatest common divisor of these l, it follows that also $e(k\psi(y)) = \text{const.}$ for almost all y. On placing $\text{const.} = e(k\alpha)$, where $0 \leqq \alpha < 1/k$, one sees that the probability that $\psi(y)$ assumes one of the values $l/k + \alpha$ is 1, i. e., that $\psi(y)$ has the distribution stated under (ii). On the other hand, if $A(n)$ is defined by (3), and $k \geqq 1$ is the greatest common divisor of all integers for which (5) holds, then the probability distribution of $\phi(x)$ over the circle $0 \leqq \phi < 1$ obviously is periodic, with primitive period $1/k$, so that also $\phi(x)$ has the distribution stated under (ii).

3. From the following theorem the missing part of the above proof of Theorem I will follow as a special case.

Theorem II. If $f(y)$ and $g(y)$ are two measurable functions of y on $0 \leqq y \leqq 1$, the first of which is integrable and the second real, then

$$\int_0^1 f(y) e(ug(y))\, dy = \int_0^1 f(y)\, dy \cdot \int_0^1 e(ug(y))\, dy \tag{7}$$

cannot hold for every real u, unless

$$\int_E f(y)\, dy = |E| \int_0^1 f(y)\, dy$$

holds for every set $E = E_\omega$ of the form

$$E = E_\omega: \qquad g(y) < \omega. \tag{8}$$

On replacing $f(y)$ by $f(y) - \int_0^1 f(y)\, dy$, one sees that Theorem II is

equivalent to the following lemma: If, for every real u,

$$\int_0^1 f(y)\, e(ug(y))\, dy = 0, \tag{9}$$

then

$$\int_E f(y)\, dy = 0, \tag{10}$$

where $E = E_\omega$ is defined by (8).

In order to prove this lemma, notice that

$$\frac{1}{\pi}\int_{-\infty}^{+\infty} u^{-1} \sin 2\pi u(\omega - g(y))\, du = \operatorname{sign}(\omega - g(y))$$

and that we may suppose that the set where $g(t) = \omega$ is a 0-set. Thus [6] the left side of (10) is, in view of (8):

$$\int_0^1 f(y)\ (\tfrac{1}{2} + \tfrac{1}{2} \operatorname{sign}(\omega - g(y)))\, dy$$

$$= \tfrac{1}{2}\int_0^1 f(y)\, dy + \tfrac{1}{2\pi}\int_{-\infty}^{+\infty} u^{-1} \int_0^1 f(y) \sin 2\pi u(\omega - g(y))\, dy du.$$

Since the right-hand side of this equation is obviously 0 by (9), the proof of (10), hence also that of Theorem II, is complete.

If $f(y)$, $g(y)$ satisfy the assumptions of Theorem II, and, in addition, $g(y)$ is monotone increasing for $0 \leqq y \leqq 1$, then every interval $0 \leqq y < a$, where $a \leqq 1$, may be obtained as a set of the form (8), as seen by placing $\omega = g(a)$. Thus, if $g(y)$ is monotone increasing on $0 \leqq y \leqq 1$, then (7) implies that $f(y) = \text{const.}$ almost everywhere (with const. $= \int_0^1 f(y)\, dy$). Hence, application of Theorem II to $g(y) = y$ and $f(y) = e(l\psi(y))$ shows that $e(l\psi(y)) = \text{const.}$ almost everywhere is the only solution of (6). This completes the proof of Theorem I.

4. It may be observed that similar independence requirements can lead to other distribution functions also.

---

[6] For the justification of the change on the order of integration, and the elimination of the condition that the set where $g(y) = \omega$ is a 0-set, *cf.*, e. g., M. Kac and H. Steinhaus, "Sur les fonctions indépendantes, III," *Studia Mathematica*, vol. 6 (1936), pp. 93-94.

In fact, suppose that for a given measurable function $f(x)$ on $(0,1)$ the angular sums $\{f(x)+f(y)\}$ and $\{f(x)-f(y)\}$ are independent in the angular sense. Denoting by A_n the Fourier-Stieltjes coefficients of the angular distribution function of $\{f(x)\}$, one has evidently from the above assumption of independence that $A_{m+n}A_{m-n}=A_m^2A_n^2$ for arbitrary integers m, n.

It is easy to see that the only solution of this equation is $A_n=q^{n^2}$ where $|q|\leqq 1$, since $|A_n|\leqq 1$, and where q is real, since $A_n=\bar{A}_{-n}$. Thus,[3] the independence condition given above leads to the Gaussian distribution which is reduced mod 1,[7] q being positive or negative according as the median of the Gaussian distribution is taken to represent the value 0 or the value $\frac{1}{2}$ of the angular variable; the case $q=\pm 1$ corresponds of course to the unit distribution.

Appendix. It is clear that Theorem II can be generalized in the following direction: If $f(t)$, $g(t)$, $h(t)$, where $0\leqq t\leqq 1$, are three measurable functions, the first of which is integrable while the other two are real, then the integral of $f(t)\,e(ug(t)+vh(t))$ over the interval $0\leqq t\leqq 1$ cannot vanish for every real u and v, unless $\int_E f(t)\,dt=0$ for every set $E=E(\omega_1,\omega_2)$ of t-values, which is defined by the pair of inequalities $g(t)<\omega_1$, $h(t)<\omega_2$, where ω_1,ω_2 are arbitrary fixed real numbers.

It follows that, if every subinterval of $(0,1)$ can be approximated by sums of sets $E=E(\omega_1,\omega_2)$, then the functions $e(u_kg(t)+v_kh(t))\,f_k(t)=e(u_kg(t)+v_kh(t))$, where (u_k,v_k) is any fixed dense sequence of points in the (u,v)-plane, form a complete set of functions with reference to the function class L on $(0,1)$.

As a particular case, one obtains the following theorem:[8]

If $g(t)$ and $h(t)$ are essentially bounded and if every subinterval of $(0,1)$ can be approximated by sums of sets $E=E(\omega_1,\omega_2)$, then the double sequence of functions $g^k(t)\,h^l(t)$, where $k,l=0,1,2,\cdots$, is a complete set in L. For instance, one can put $g(t)=\sin 2\pi t$ and $h(t)=\sin 4\pi t$.

THE JOHNS HOPKINS UNIVERSITY.

---

[7] According to a remark of Aurel Wintner, the orthogonal trigonometrical polynomials investigated by G. Szegö ("Ein Beitrag zur Theorie der Thetafunktionen," *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, vol. 19 [1926], pp. 242-252), may be interpreted as the angular analogues to the Hermitian polynomials, which belong to the linear Gaussian distribution.

[8] This is a particular case of complete systems considered by B. v. Sz. Nagy, "Zur Theorie der Charakteren abelscher Gruppen," *Mathematische Annalen*, vol. 114 (1937), pp. 373-384.

ON OSCILLATIONS OF REAL SEQUENCES AND OF THEIR TRANSFORMS BY SQUARE MATRICES.* [1]

By RALPH PALMER AGNEW.[2]

1. Introduction. Let A denote simultaneously a matrix a_{nk} $(n, k = 1, 2, 3, \cdots)$ of complex constants, and the transformation

$$A \qquad y_n = \sum_{k=1}^{\infty} a_{nk} x_k$$

which associates with each given sequence x_n of complex numbers (in the range of application of A) a transform y_n. The *oscillation* of a sequence s_n of complex numbers is defined by

$$(1.01) \qquad \Omega(s) = \limsup_{m,n\to\infty} | s_m - s_n |.$$

Observe that $\Omega(s)$ is 0, finite and > 0, or $+\infty$ according as s_n is convergent, bounded and divergent, or unbounded. For a bounded divergent sequence s_n, the oscillation $\Omega(s)$ is the maximum of all distances between pairs of limit points of the sequence. If a sequence x_n lies in the range of application of A, there arises the question whether $\Omega(y) \leqq \Omega(x)$. In case x_n is unbounded so that $\Omega(x) = +\infty$, the inequality in which we are interested becomes trivial. Hence we confine our attention to bounded sequences.

The main problem of this paper is that of characterizing the matrices A having the property that $\Omega(y) \leqq \Omega(x)$ for each real bounded divergent sequence x_n.

Two problems related to the above problem are solved by the following two theorems.[3]

THEOREM 1.1. *In order that a regular transformation A may be such that $\Omega(y) \leqq \Omega(x)$ for each bounded sequence x_n, the condition*

* Received March 16, 1938.

[1] Presented to the American Mathematical Society April 15, 1938.

[2] The author is indebted to Professor W. A. Hurwitz for numerous helpful suggestions and criticisms.

[3] These theorems are proved for triangular matrices (matrices for which $a_{nk} = 0$ when $k > n$) by W. A. Hurwitz, "The oscillation of a sequence," *American Journal of Mathematics*, vol. 52 (1930), pp. 611-616. The extension to square matrices is given by R. P. Agnew, "The effects of general regular transformations on oscillations of sequences of functions," *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 411-424.

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} | a_{nk} | = 1 \tag{1.11}$$

is necessary and sufficient.

THEOREM 1.2. *In order that a regular transformation A with real elements may be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n,* (1.11) *is necessary and sufficient.*

It is the rather unexpected circumstance (see § 5) that Theorem 1.2 fails to hold when the phrase "with real elements" is removed which caused the author to seek in the first place a characterization of regular methods A such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n, and in the second place to propose the more general problem previously mentioned.

2. Conservative matrices. A matrix A is called *conservative* (convergence-preserving) if the A-transform y_n of each convergent complex sequence x_n is convergent. It follows from this definition that A is conservative if and only if $\Omega(y) = 0$ for each complex convergent sequence x_n. Necessary and sufficient conditions that A be conservative are[4] that there exist a real constant M, a set $\alpha_1, \alpha_2, \cdots$ of complex constants, and a complex constant α such that

$$\sum_{k=1}^{\infty} | a_{nk} | \leqq M \qquad (n = 1, 2, \cdots) \tag{2.01}$$

$$\lim_{n\to\infty} a_{nk} = \alpha_k \qquad (k = 1, 2, \cdots) \tag{2.02}$$

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{nk} = \alpha. \tag{2.03}$$

Our first step in the solution of our problem is to prove the following lemma.

LEMMA 2.1. *If A is such that $\Omega(y) \leqq \Omega(x)$ for each real bounded divergent sequence x_n, then A is conservative.*

Let x_n be a real convergent sequence. We show that $\Omega(y) = 0$. Let $\epsilon > 0$. Since x_n is real and convergent, the sequences x_n' and x_n'' defined by

$$x_n' = (-1)^n\epsilon, \qquad x_n'' = x_n - (-1)^n\epsilon \qquad (n = 1, 2, \cdots) \tag{2.11}$$

[4] T. Kojima, "On generalized Toeplitz's theorems on limit and their applications," *Tohoku Mathematical Journal*, vol. 12 (1917), pp. 291-326; I. Schur, "Über lineare Transformationen in der Theorie der unendlichen Reihen," *Crelle Journal*, vol. 151 (1921), pp. 79-111.

are both real and divergent, and we see that

$$\Omega(x') = \Omega(x'') = 2\epsilon. \tag{2.12}$$

Hence our hypothesis guarantees existence of the A-transforms y_n' and y_n'' of x_n' and x_n'', and the inequalities

$$\Omega(y') \leqq 2\epsilon, \qquad \Omega(y'') \leqq 2\epsilon. \tag{2.13}$$

Since $x_n = x_n' + x_n''$ and A is linear, the transform y_n of x_n must exist and $y_n = y_n' + y_n''$. Therefore

$$\Omega(y) \leqq \Omega(y') + \Omega(y'') \leqq 4\epsilon. \tag{2.14}$$

As ϵ is an arbitrary positive number, $\Omega(y) = 0$. We have shown that the A-transform of each real convergent sequence is convergent. Since A is linear, it follows in turn that the A-transform of each pure imaginary and each complex convergent sequence is convergent. Hence, by definition, A is conservative and Lemma 2.1 is proved.

THEOREM 2.2. *In order that A may be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded divergent sequence x_n, it is necessary and sufficient that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n.*

Necessity follows from Lemma 2.1, and sufficiency is trivial.

If a matrix A is conservative, it follows from (2.01) that $\sum_{k=1}^{p} | a_{nk} | \leqq M$ for each $n, p = 1, 2, \cdots$ and using (2.02) gives $\sum_{k=1}^{p} | \alpha_k | \leqq M$ for each $p = 1, 2, \cdots$. Hence $\sum_{k=1}^{\infty} | \alpha_k | \leqq M$. The matrix A can be represented as the sum of two matrices A' and A'' defined by

$$a'_{nk} = \alpha_k; \qquad a''_{nk} = a_{nk} - \alpha_k \qquad (n, k = 1, 2, \cdots).$$

The A' and A'' transforms of each bounded sequence x_n exist and are given by

$$y_n' = \sum_{k=1}^{\infty} \alpha_k x_k, \qquad y_n'' = y_n - \sum_{k=1}^{\infty} \alpha_k x_k.$$

Hence $\Omega(y') = 0$ and $\Omega(y'') = \Omega(y)$. Accordingly each of the two inequalities $\Omega(y) \leqq \Omega(x)$ and $\Omega(y'') \leqq \Omega(x)$ implies the other.

The matrix A' is conservative and columnwise-constant. The matrix A'' is such that $\sum_{k=1}^{\infty} | a''_{nk} | \leqq M + \sum_{k=1}^{\infty} | \alpha_k |$, $\lim_{n\to\infty} a''_{nk} = 0$ for each k, and $\lim_{n\to\infty} \sum_{k=1}^{\infty} a''_{nk}$ exists. The last limit is, in fact, $\rho \equiv \alpha - \sum_{k=1}^{\infty} \alpha_k$.

Thus the matrix A'' belongs to the class of *multiplicative* matrices,[5] i. e. matrices A for which

$$\sum_{k=1}^{\infty} |\, a_{nk} \,| < M \qquad (n = 1, 2, \cdots) \tag{2.21}$$

$$\lim_{n\to\infty} a_{nk} = 0 \qquad (k = 1, 2, \cdots) \tag{2.22}$$

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} a_{nk} = \rho. \tag{2.23}$$

The preceding discussion and definitions furnish the basis for proof of the following theorem

THEOREM 2.3. *In order that A may be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n, it is necessary and sufficient that A be the sum of a conservative columnwise-constant matrix A' and a multiplicative matrix A'' such that $\Omega(y'') \leqq \Omega(x)$ for each real bounded sequence x_n.*

On account of Theorems 2.2 and 2.3, our problem reduces to that of characterizing the multiplicative matrices A having the property that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n; henceforth we confine our attention to multiplicative matrices.

3. Lemmas involving multiplicative matrices. In this section, we give a number of lemmas each one of which applies to the class of multiplicative matrices. In § 7, we shall use the fact that each lemma after the first applies to a more general class of matrices.

LEMMA 3.1. *If A is a multiplicative matrix, then a necessary and sufficient condition that*

$$\Omega(y) \leqq \Omega(x) \qquad x_n \text{ real and bounded} \tag{3.11}$$

is that

$$\Omega(y) \leqq 2 \qquad -1 \leqq x_n \leqq 1. \tag{3.12}$$

The meaning of (3.11) is that $\Omega(y)$ must be less than or equal to $\Omega(x)$ for *each* real bounded sequence x_n, while (3.12) means that $\Omega(y)$ must be less than or equal to 2 for *each* real sequence x_n whose elements satisfy the condition $-1 \leqq x_n \leqq 1$.

Necessity is obvious. To prove sufficiency, we assume that (2.21), (2.22), and (2.23) hold and (3.11) fails, and show that (3.12) must fail.

---

[5] The reason for calling a matrix A satisfying (2.21), (2.22) and (2.23) multiplicative is that if $x_n \to x$, then the A-transform y_n of x_n exists and $y_n \to \rho x$, i. e. the limit of y_n is obtained by multiplying the limit of x_n by ρ. A multiplicative matrix A is null if the multiplier ρ is 0, and is regular if $\rho = 1$.

Since (3.11) fails, there is a real bounded sequence x_n such that $\Omega(y) > \Omega(x)$. Since A is multiplicative, x_n must be divergent so that

$$\Omega(x) = \Lambda - \lambda > 0$$

where

$$\lambda = \liminf_{n\to\infty} x_n, \qquad \Lambda = \limsup_{n\to\infty} x_n.$$

Let the sequence x'_n be defined by

$$x'_n = (2x_n - \Lambda - \lambda)/(\Lambda - \lambda) \qquad (n = 1, 2, \cdots).$$

Then $\liminf x'_n = -1$, $\limsup x'_n = 1$, and $\Omega(x') = 2$. Letting y'_n denote the transform of x'_n, we have

$$y'_m - y'_n = \sum_{k=1}^{\infty} a_{mk}\left(\frac{2x_k}{\Lambda - \lambda} - \frac{\Lambda + \lambda}{\Lambda - \lambda}\right) - \sum_{k=1}^{\infty} a_{nk}\left(\frac{2x_k}{\Lambda - \lambda} - \frac{\Lambda + \lambda}{\Lambda - \lambda}\right)$$

and using (2.23) we obtain

$$y'_m - y'_n = \epsilon_{mn} + 2(y_m - y_n)/(\Lambda - \lambda),$$

where $\epsilon_{mn} \to 0$ as $m, n \to \infty$. Hence

$$\Omega(y') = 2\Omega(y)/(\Lambda - \lambda) > 2\Omega(x)/(\Lambda - \lambda) = 2. \tag{3.13}$$

Finally, let $x''_n = -1$, x'_n, or 1 according as $x'_n \leqq -1$, $-1 < x'_n < 1$, or $x'_n \geqq 1$. Using the equalities $\liminf x'_n = -1$ and $\limsup x'_n = 1$, we can show that $x''_n - x'_n \to 0$ as $n \to \infty$. Letting y''_n denote the transform of x''_n, we find

$$y''_n - y'_n = \sum_{k=1}^{\infty} a_{nk}(x''_k - x'_k). \tag{3.14}$$

But the two properties (2.21) and (2.22) of A are sufficient to ensure that if $s_n \to 0$, then $\sum_{k=1}^{\infty} a_{nk}s_k \to 0$. Hence $y''_n - y'_n \to 0$, and it follows that $\Omega(y'') = \Omega(y')$. Therefore $\Omega(y'') > 2$ while $-1 \leqq x''_n \leqq 1$ for all n. This violates (3.12) and Lemma 3.1 is proved.

LEMMA 3.2. *If A satisfies (2.21) and (2.22), then a necessary and sufficient condition that*

$$\Omega(y) \leqq 2 \qquad -1 \leqq x_n \leqq 1 \tag{3.21}$$

is that

$$\limsup_{n\to\infty} |y_n| \leqq 1 \qquad -1 \leqq x_n \leqq 1. \tag{3.22}$$

To prove sufficiency, we observe that if (3.22) holds and $-1 \leqq x_n \leqq 1$, then the inequality $|y_m - y_n| \leqq |y_m| + |y_n|$ and (3.22) yield (3.21).

To prove necessity, we assume that (3.22) fails and prove that (3.21) must fail. Let x_n be a sequence with $-1 \leqq x_n \leqq 1$ for which $\limsup |\, y_n \,| > 1$, and let a limit point ζ of y_n be selected such that $|\, \zeta \,| > 1$. Let $k_0 = 0$, let n_1 be an arbitrary positive integer, and choose $k_1 > k_0$ such that

$$\sum_{k=k_1+1}^{\infty} |\, a_{n_1 k} \,| < 1.$$

Then for each sufficiently great index $n_2 > n_1$, we have

$$\sum_{k=1}^{k_1} |\, a_{n_2 k} \,| < \tfrac{1}{2},$$

and when n_2 is fixed we can choose $k_2 > k_1$ such that

$$\sum_{k=k_2+1}^{\infty} |\, a_{n_2 k} \,| < \tfrac{1}{2}.$$

Continuing in this way, we obtain sequences

$$0 = k_0 < k_1 < k_2 < \cdots ; \quad n_1 < n_2 < n_3 < \cdots \tag{3.23}$$

of indices such that

$$\sum_{k=1}^{k_{p-1}} |\, a_{n_p k} \,| < 1/p, \qquad \sum_{k=k_p+1}^{\infty} |\, a_{n_p k} \,| < 1/p \qquad (p = 1, 2, 3, \cdots). \tag{3.24}$$

Arbitrariness of choice of the indices (3.23) enables us to choose them such that we have, in addition to (3.24), $\lim y(n_p) = \zeta$. We sometimes use $y(j)$ to denote y_j when $j = n_p$. Using (3.24) and the inequality $-1 \leqq x_n \leqq 1$, we find that

$$y_{n_p} = 2\epsilon_p + \sum_{k=k_{p-1}+1}^{k_p} a_{n_p k} x_k$$

where $|\, \epsilon_p \,| < 1/p$. Now let a sequence x'_n be defined as follows: for each index k, choose the index p such that $k_{p-1} < k \leqq k_p$ and put $x'_k = (-1)^p x_k$. Then $-1 \leqq x'_n \leqq 1$ and the transform y'_n of x'_n is such that

$$y'_{n_p} = 4\epsilon_p + (-1)^p y_{n_p}.$$

Therefore both ζ and $-\zeta$ are limit points of y'_n so that $\Omega(y') \geqq 2\,|\, \zeta \,| > 2$. Thus (3.21) fails and Lemma 3.2 is proved.

LEMMA 3.3. *If A satisfies* (2.21) *and* (2.22), *then a necessary and sufficient condition that*

$$\limsup_{n\to\infty} |\, y_n \,| \leqq 1 \qquad -1 \leqq x_n \leqq 1 \tag{3.31}$$

is that

$$\limsup_{n\to\infty} |\, y_n \,| \leqq 1 \qquad x_n = \pm 1. \tag{3.32}$$

Necessity is trivial. To prove sufficiency, let x_n be a sequence with $-1 \leqq x_n \leqq 1$ for which

$$\limsup_{n\to\infty} \left| \sum_{k=1}^{\infty} a_{nk}x_k \right| > \Lambda > 1. \tag{3.33}$$

We show that (3.32) must fail. It follows from (3.33) and properties of A that we can determine a system (3.23) of indices for which (3.24) holds and

$$\left| \sum_{k=k_{p-1}+1}^{k_p} a_{n_p k}x_k \right| > \Lambda - 2/p \qquad (p = 1, 2, \cdots). \tag{3.34}$$

For each $p = 1, 2, \cdots$, it is possible to define $x'_k = +1$ or -1 over the range $k_{p-1} + 1 \leqq k \leqq k_p$ in such a way that the left member of (3.34) will not be decreased when x_k is replaced by x'_k. This may be seen by making iterated use of the elementary fact that if z_1 and z_2 are complex numbers, then the maximum value of $| z_1 + xz_2 |$ over the interval $-1 \leqq x \leqq 1$ is attained when $x = -1$ or $x = 1$. We thus obtain a sequence x'_n of which each element is either $+1$ or -1, and can show that the transform y'_n of x'_n is such that $\limsup | y'_n | \geqq \Lambda > 1$. This contradicts (3.32), and completes the proof of Lemma 3.3.

LEMMA 3.4. *If A satisfies (2.21) and (2.22), then a necessary and sufficient condition that*

$$\limsup_{n\to\infty} | y_n | \leqq 1 \qquad x_n = \pm 1 \tag{3.41}$$

is that

$$\limsup_{n\to\infty} \text{l. u. b.}_{U_{nk}=\pm 1} \left| \sum_{k=1}^{\infty} a_{nk}U_{nk} \right| \leqq 1. \tag{3.42}$$

The condition (3.42) is equivalent to the condition: for each $\epsilon > 0$ there is an index N such that

$$\left| \sum_{k=1}^{\infty} a_{nk}x_k \right| < 1 + \epsilon \qquad n \geqq N;\ x_k = \pm 1. \tag{3.43}$$

Thus, while (3.41) asserts that $\limsup | y_n | \leqq 1$ for each sequence x_n of which the elements are either ± 1, (3.42) asserts that $\limsup | y_n | \leqq 1$ *uniformly* for all sequences x_n of which the elements are either ± 1.

Sufficiency is obvious. To prove necessity, suppose (3.42) fails. Then we can choose $\delta > 0$ and a system (3.23) of indices such that (3.24) holds and also

$$\text{l. u. b.}_{U_{nk}=\pm 1} \left| \sum_{k=1}^{\infty} a_{nk}U_{nk} \right| > 1 + \delta \qquad n = n_1, n_2, \cdots.$$

For each index p, let $x_n^{(p)}$ be a sequence such that $x_n^{(p)} = \pm 1$ for each n and

$$|\sum_{k=1}^{\infty} a_{n_p k} x_k^{(p)}| > 1 + \delta.$$

Let $x'_k = x_k^{(p)}$ when $k_{p-1} < k \leqq k_p$. Then $x'_n = \pm 1$ for each n. But the transform y'_n of x'_n is such that

$$\begin{aligned} |y'_{n_p}| &\geqq -\sum_{k=1}^{n_{p-1}} |a_{n_p k}| + |\sum_{k=n_{p-1}+1}^{n_p} a_{n_p k} x'_k| - \sum_{k=n_p+1}^{\infty} |a_{n_p k}| \\ &\geqq -2/p + |\sum_{k=n_{p-1}+1}^{n_p} a_{n_p k} x_k^{(p)}| \\ &\geqq -4/p + |\sum_{k=1}^{\infty} a_{n_p k} x_k^{(p)}| > 1 + \delta - 4/p. \end{aligned}$$

Hence lim sup $|y'_n| \geqq 1 + \delta > 1$ and this contradiction of (3. 41) establishes Lemma 3. 4.

LEMMA 3. 5. *If $a_1, a_2, \cdots$ are complex numbers such that*

$$\sum_{k=1}^{\infty} |a_{nk}| < \infty, \tag{3. 51}$$

then

$$\text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\mathcal{R}(e^{i\theta} a_k)| = \text{l. u. b.}_{U_k = \pm 1} |\sum_{k=1}^{\infty} a_k U_k|. \tag{3. 52}$$

We use $\mathcal{R}(z)$ and $\mathcal{I}(z)$ to denote the real and pure imaginary parts s and it of a complex number $z = s + it$. The sum $\Sigma |\mathcal{R}(e^{i\theta} a_k)|$ can be described as the sum of the distances of the points a_k from the line l_θ, through the origin of the complex plane, whose normal makes an angle $-\theta$ with the positive real axis; and the left member of (3. 52) is the least upper bound for all lines l_θ of this sum.

To prove Lemma 3. 5, let, for each θ in the interval $-\pi < \theta \leqq \pi$, $U_k^{(\theta)}$ be $+1$ or -1 according as $\mathcal{R}(e^{i\theta} a_k) \geqq 0$ or $\mathcal{R}(e^{i\theta} a_k) < 0$. Then for each θ

$$\begin{aligned} \sum_{k=1}^{\infty} |\mathcal{R}(e^{i\theta} a_k)| &= \sum_{k=1}^{\infty} \mathcal{R}(e^{i\theta} a_k U_k^{(\theta)}) \\ &= \mathcal{R}(\sum_{k=1}^{\infty} e^{i\theta} a_k U_k^{(\theta)}) \leqq |\sum_{k=1}^{\infty} e^{i\theta} a_k U_k^{(\theta)}| = |\sum_{k=1}^{\infty} a_k U_k^{(\theta)}|. \end{aligned} \tag{3. 53}$$

Using this inequality and the fact that $U_k^{(\theta)} = \pm 1$ for each θ and k, we obtain

$$\text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\mathcal{R}(e^{i\theta} a_k)| \leqq \text{l. u. b.}_{-\pi < \theta \leqq \pi} |\sum_{k=1}^{\infty} a_k U_k^{(\theta)}| \leqq \text{l. u. b.}_{U_k = \pm 1} |\sum_{k=1}^{\infty} a_k U_k|. \tag{3. 54}$$

Let K denote the right member of (3. 52) and let $\epsilon > 0$. Choose $\tilde{U}_1, \tilde{U}_2, \cdots$

such that $\tilde{U}_k = \pm 1$ for each k and $\sum_{k=1}^{\infty} a_k \tilde{U}_k = \zeta$ where $|\zeta| > K - \epsilon$. Let ρ and ϕ be determined such that $\rho \geqq 0$, $-\pi < \phi \leqq \pi$, and $\zeta = \rho e^{-i\phi}$. Then

$$\sum_{k=1}^{\infty} e^{i\phi} a_k \tilde{U}_k = \rho$$

is real. We then have

$$\begin{aligned} (3.55) \qquad \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\phi} a_k)| &= \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\phi} a_k \tilde{U}_k)| \\ &\geqq |\sum_{k=1}^{\infty} \mathcal{R}(e^{i\phi} a_k \tilde{U}_k)| = \rho = |\zeta| > K - \epsilon. \end{aligned}$$

It follows from (3.55), arbitrariness of $\epsilon > 0$, and the definition of K that

$$(3.56) \qquad \text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\theta} a_k)| \geqq \text{l. u. b.}_{U_k = \pm 1} |\sum_{k=1}^{\infty} a_k U_k|.$$

It follows from (3.54) and (3.56) that all three members of (3.54) are equal and Lemma 3.5 is proved.

From Lemma 3.5 we obtain immediately

LEMMA 3.6. *If*

$$(3.61) \qquad \sum_{k=1}^{\infty} |\, a_{nk}| < \infty \qquad (n = 1, 2, \cdots)$$

then a necessary and sufficient condition that

$$(3.62) \qquad \limsup_{n\to\infty} \text{l. u. b.}_{U_{nk} = \pm 1} |\sum_{k=1}^{\infty} a_{nk} U_{nk}| \leqq 1$$

is that

$$(3.63) \qquad \limsup_{n\to\infty} \text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\theta} a_{nk})| \leqq 1.$$

LEMMA 3.7. *If*

$$\sum_{k=1}^{\infty} |\, a_{nk}| < M \qquad (n = 1, 2, \cdots),$$

then a necessary and sufficient condition that

$$(3.71) \qquad \limsup_{n\to\infty} \text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\theta} a_{nk})| \leqq 1$$

is that

$$(3.72) \qquad \limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\theta} a_{nk})| \leqq 1 \qquad -\pi < \theta \leqq \pi.$$

Necessity is trivial. To prove sufficiency, let ν be an arbitrary but fixed

positive integer, and let $\phi_1, \phi_2, \cdots, \phi_{2\nu}$ denote in order the numbers $(-\nu+1)\pi/\nu$, $(-\nu+2)\pi/\nu, \cdots, \nu\pi/\nu$. Choose an index N such that

$$\sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\phi_p} a_{nk})\,| < 1 + \nu^{-1} \qquad (p = 1, 2, \cdots, 2\nu;\ n > N). \tag{3.73}$$

If θ is fixed in the interval $-\pi < \theta \leqq \pi$, we can choose an index p such that $|\,\theta - \phi_p\,| < \pi/\nu$ and obtain for each n the inequality

$$\begin{aligned}
&\sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\theta} a_{nk})\,| - \sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\phi_p} a_{nk})\,| \\
&\leqq \sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\theta} a_{nk}) - \mathcal{R}\,(e^{i\phi_p} a_{nk})\,| = \sum_{k=1}^{\infty} |\, \mathcal{R}\,([e^{i\theta} - e^{i\phi_p}] a_{nk})\,| \\
&\leqq \sum_{k=1}^{\infty} |\, e^{i\theta} - e^{i\phi_p}\,|\, |\, a_{nk}\,| \leqq M\,|\, 1 - e^{i(\phi_p - \theta)}\,| < M\,|\, 1 - e^{i\pi/\nu}\,|.
\end{aligned}$$

It follows from this inequality and (3.73) that when $-\pi < \theta \leqq \pi$ and $n > N$

$$\sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\theta} a_{nk})\,| < 1 + \nu^{-1} + M\,|\, 1 - e^{i\pi/\nu}\,|. \tag{3.74}$$

Hence the left member of (3.71) is less than or equal to the right member of (3.74) and we can let $\nu \to \infty$ to obtain (3.71). Thus Lemma 3.7 is proved.

4. A theorem on multiplicative matrices. The following theorem gives several necessary and sufficient conditions that a multiplicative matrix A be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n. The author gives the whole set because it appears that no one of the criteria is always more convenient (or always less convenient) than any other one.

THEOREM 4.1. *If A is a multiplicative matrix with multiplier ρ, then each one of the following conditions implies each other one:*

$$\Omega(y) \leqq \Omega(x) \qquad x_n \text{ real and bounded} \tag{4.11}$$

$$\Omega(y) \leqq 2 \qquad -1 \leqq x_n \leqq 1 \tag{4.12}$$

$$\limsup_{n\to\infty} |\, \sum_{k=1}^{\infty} a_{nk} x_k\,| \leqq 1 \qquad -1 \leqq x_n \leqq 1 \tag{4.13}$$

$$\limsup_{n\to\infty} |\, \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} t_k\,| \leqq \tfrac{1}{2} \qquad 0 \leqq t_n \leqq 1. \tag{4.131}$$

$$\limsup_{n\to\infty} |\, \sum_{k=1}^{\infty} a_{nk} U_k\,| \leqq 1 \qquad U_k = \pm 1 \tag{4.14}$$

$$\limsup_{n\to\infty} |\, \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} V_k\,| \leqq \tfrac{1}{2} \qquad V_k = 0 \text{ or } 1 \tag{4.141}$$

$$\limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, \Re\,(e^{i\theta} a_{nk})\,| \leqq 1 \qquad -\pi < \theta \leqq \pi \tag{4.142}$$

$$\limsup_{n\to\infty} \text{l. u. b.}_{U_{nk}=\pm 1} \Big| \sum_{k=1}^{\infty} a_{nk} U_{nk} \Big| \leqq 1 \tag{4.15}$$

$$\limsup_{n\to\infty} \text{l. u. b.}_{V_{nk}=0 \text{ or } 1} \Big| \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} V_{nk} \Big| \leqq \tfrac{1}{2} \tag{4.151}$$

$$\limsup_{n\to\infty} \text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, \Re\,(e^{i\theta} a_{nk})\,| \leqq 1. \tag{4.152}$$

It follows from the lemmas of § 3 that each one of the conditions (4. 11), (4. 12), (4. 13), (4. 14), (4. 15), (4. 152) and (4. 142) implies each other one. We show that (4. 15) and (4. 151) imply each other. For each fixed n, we have

$$\begin{aligned} \text{l. u. b.}_{U_{nk}=\pm 1} \Big| \sum_{k=1}^{\infty} a_{nk} U_{nk} \Big| &= \text{l. u. b.}_{W_{nk}-1=\pm 1} \Big| \sum_{k=1}^{\infty} a_{nk} - \sum_{k=1}^{\infty} a_{nk} W_{nk} \Big| \\ &= 2 \text{ l. u. b.}_{V_{nk}=0 \text{ or } 1} \Big| \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} V_{nk} - \tfrac{1}{2}\Big(\rho - \sum_{k=1}^{\infty} a_{nk}\Big) \Big| \\ &= \epsilon_n + 2 \text{ l. u. b.}_{V_{nk}=0 \text{ or } 1} \Big| \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} V_{nk} \Big| \end{aligned}$$

where, in consequence of (2. 23), $\epsilon_n \to 0$ as $n \to \infty$. Taking superior limits as $n \to \infty$, we see that the left member of (4. 15) is double the left member of (4. 151). Hence each of (4. 15) and (4. 151) implies the other. It can be shown similarly that each of (4. 13) and (4. 131) implies the other, and that each of (4. 14) and (4. 141) implies the other. This completes the proof of Theorem 4. 1.

In any case where $\sum_{k=1}^{\infty} |\, a_{nk}\,| < \infty$ for each n, it is possible to write the condition (4. 152) in several equivalent forms. Since, for each n,

$$f_n(\theta) = \sum_{k=1}^{\infty} |\, \Re\,(e^{i\theta} a_{nk})\,| \tag{4.16}$$

is a continuous function of θ of period π, and

$$f_n(\theta + \pi/2) = \sum_{k=1}^{\infty} |\, \Im\,(e^{i\theta} a_{nk})\,|,$$

we obtain a condition equivalent to (4. 152) by making any one or more of the following changes in the left member of (4. 152): (a) replace "l. u. b." by "max"; (b) replace the interval $-\pi < \theta \leqq \pi$ over which the least upper bound (or maximum) is taken by any interval of length π; (c) replace $\Re$ by $\Im$; and (d) replace θ by $-\theta$.

If (4.152) holds, we can set $\theta = 0$ and $\theta = \pi/2$ to obtain

$$(4.17) \qquad \limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, \mathcal{R}(a_{nk})\,| \leqq 1, \quad \limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, \mathcal{I}(a_{nk})\,| \leqq 1.$$

Thus Theorem 4.1 implies that if A is multiplicative and (4.11) holds, then (4.17) holds and hence also

$$(4.18) \qquad \limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, a_{nk}\,| \leqq 2.$$

In § 6 we obtain a necessary condition stronger than (4.18).

In case A is a real matrix, the condition (4.152) becomes

$$\limsup_{n\to\infty} \text{ l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, a_{nk}\,| \;|\, \mathcal{R}(e^{i\theta})\,| \leqq 1$$

which simplifies to

$$(4.19) \qquad \limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, a_{nk}\,| \leqq 1.$$

Thus (4.19) is a very simple necessary and sufficient condition that a real multiplicative matrix A may be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n.

In case $|\,\rho\,| = 1$, it is possible to modify conditions (4.15), (4.151), and (4.152). Since the class of multiplicative matrices for which $|\,\rho\,| = 1$ contains as an important subclass the class of all regular matrices, we give our modifications in the following theorem.

THEOREM 4.2. *If A is a multiplicative matrix with multiplier ρ and $|\,\rho\,| = 1$, then each one of the following conditions implies each other one:*

$$(4.21) \qquad \Omega(y) \leqq \Omega(x) \qquad x_n \text{ real and bounded}$$

$$(4.22) \qquad \lim_{n\to\infty} \text{ l. u. b.}_{U_{nk}=\pm 1} \Big|\, \sum_{k=1}^{\infty} a_{nk} U_{nk}\,\Big| = 1$$

$$(4.23) \qquad \lim_{n\to\infty} \text{ l. u. b.}_{V_{nk}=0 \text{ or } 1} \Big|\, \tfrac{1}{2}\rho - \sum_{k=1}^{\infty} a_{nk} V_{nk}\,\Big| = \tfrac{1}{2}$$

$$(4.24) \qquad \lim_{n\to\infty} \text{ l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} |\, \mathcal{R}(e^{i\theta} a_{nk})\,| = 1.$$

It is evident that (4.22) implies (4.15) and hence, by Theorem 4.1, (4.21). The hypothesis $|\,\rho\,| = 1$ gives

$$(4.25) \qquad \lim_{n\to\infty} \Big|\, \sum_{k=1}^{\infty} a_{nk}\,\Big| = |\,\rho\,| = 1.$$

Using (4.25) and the inequality

$$\text{l. u. b.}_{U_{nk}=\pm 1} \left| \sum_{k=1}^{\infty} a_{nk} U_{nk} \right| \geqq \left| \sum_{k=1}^{\infty} a_{nk} \right| \qquad (n = 1, 2, \cdots)$$

we obtain

$$\liminf_{n\to\infty} \text{l. u. b.}_{U_{nk}=\pm 1} \left| \sum_{k=1}^{\infty} a_{nk} U_{nk} \right| \geqq 1. \tag{4.26}$$

By Theorem 4. 1, (4. 21) implies (4. 15). But (4. 15) and (4. 26) together imply (4. 22). Thus (4. 21) and (4. 22) are equivalent. Proof of equivalence of (4. 22) and (4. 23) is analogous to that of (4. 15) and (4. 151). Finally, equivalence of (4. 22) and (4. 24) follows from Lemma 3. 5, and Theorem 4. 2 is proved.

5. Some special matrices. The transformation

$$y_n = a_n x_n + (\rho - a_n + \epsilon_n) x_{n+1}, \tag{5.11}$$

where a_n and ϵ_n are complex sequences, a_n is bounded, $\epsilon_n \to 0$, and ρ is a complex constant, is multiplicative; and each multiplicative transformation of the form $y_n = a_n x_n + b_n x_{n+1}$ can be thrown into the form (5. 11). For the transformation (5. 11), the values of $\left| \sum_{k=1}^{\infty} a_{nk} U_{nk} \right|$ obtained by giving values ± 1 to U_{nk} are $| \rho + \epsilon_n |$ and $| \rho - 2a_n + \epsilon_n |$. Hence necessary and sufficient conditions that (4. 15) hold are

$$| \rho | \leqq 1; \qquad \limsup_{n\to\infty} | \tfrac{1}{2}\rho - a_n | \leqq \tfrac{1}{2}. \tag{5.12}$$

Applying Theorem 4. 1, we see that (5. 11) is such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n if and only if (5. 12) holds. In case (5. 11) is regular (i. e. $\rho = 1$), (5. 12) is equivalent to the condition that all limit points of the sequence a_n lie within or on the circle with center at the point $z = \frac{1}{2}$ and radius $\frac{1}{2}$.

In particular, the transformation

$$y_n = \frac{1+i}{2} x_n + \frac{1-i}{2} x_{n+1} \tag{5.13}$$

is regular and such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence. For this transformation,

$$\sum_{k=1}^{\infty} | a_{nk} | = \sqrt{2}, \qquad (n = 1, 2, \cdots). \tag{5.14}$$

Observe that also

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} | \Re(a_{nk}) | = \lim_{n\to\infty} \sum_{k=1}^{\infty} | \Im(a_{nk}) | = 1 \tag{5.15}$$

and hence that the necessary conditions (4. 17) are the best possible ones of their type. It follows from (5. 14) and Theorem 1. 1 that the transformation

(5. 13) does not have the property that $\Omega(y) \leqq \Omega(x)$ for each bounded complex sequence x_n. In fact, Theorem 1. 1 implies that the regular transformation

$$y_n = a_n x_n + (1 - a_n + \epsilon_n) x_{n+1} \tag{5. 16}$$

is such that $\Omega(y) \leqq \Omega(x)$ for each complex sequence x_n if and only if

$$\lim_{n\to\infty} [\,|\, a_n \,|\, + \,|\, 1 - a_n + \epsilon_n \,|\,] = 1, \tag{5. 17}$$

i. e. if and only if the limit points of the sequence a_n all lie on the closed real interval $0 \leqq z \leqq 1$.

The transformation

$$y_n = \tfrac{1}{2} x_n + \tfrac{1}{2}\omega x_{n+1} + \tfrac{1}{2}\omega^2 x_{n+2} \tag{5. 18}$$

where ω is a primitive cube root of unity, is an example of a multiplicative transformation with multiplier 0 for which (4. 15) and hence (4. 11) hold. Here

$$\sum_{k=1}^{\infty} |\, a_{nk} \,| = 3/2 \qquad (n = 1, 2, \cdots). \tag{5. 19}$$

By changing the signs of the coefficients of x_{n+1} and x_{n+2} in (5. 18), we obtain a regular matrix satisfying (4. 11) for which (5. 19) holds.

We now give a more complicated example which is of interest in connection with a necessary condition which we shall derive in the next section. Let the matrix A be defined as follows: For each $n = 1, 2, 3, \cdots$, let

$$a_{nk} = \rho_{nk} e^{i\theta_{nk}} \qquad (k = 1, 2, \cdots) \tag{5. 21}$$

where

$$\begin{aligned} \rho_{nk} &= \sin(\pi/2n) && (k = 1, 2, \cdots, n) \\ &= 0 && k > n \end{aligned} \tag{5. 22}$$

and

$$\theta_{nk} = -\pi/2 + (2k - 1)\pi/2n \qquad (k = 1, 2, \cdots). \tag{5. 23}$$

The matrix A is triangular. For each $n = 1, 2, \cdots$, the n non-vanishing elements $a_{n1}, a_{n2}, \cdots, a_{nn}$ of the n-th row of A are determined in the complex plane as the intersections of the circle $|\, z \,| = \sin(\pi/2n)$ and the rays which bisect the n equal wedges with vertex at the origin which divide the right half-plane into n equal parts. If we put $\omega_n = e^{\pi i/n}$ so that $\omega_n{}^n = -1$, we can show that for each $n = 1, 2, \cdots$

$$\begin{aligned} a_{nk} &= \tfrac{1}{2}(\omega_n{}^{k-1} - \omega_n{}^k) && 1 \leqq k \leqq n \\ &= 0 && k > n. \end{aligned} \tag{5. 24}$$

It is obvious from (5. 21) and (5. 22) that

$$\sum_{k=1}^{\infty} |\, a_{nk} \,| = n \sin(\pi/2n) < \pi/2 \qquad (n = 1, 2, \cdots) \tag{5. 25}$$

and hence that

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} |\, a_{nk} \,| = \pi/2. \tag{5.26}$$

Either (5.21) or (5.24) implies readily that

$$\lim_{n\to\infty} a_{nk} = 0 \qquad (k = 1, 2, \cdots), \tag{5.27}$$

and using (5.24) we see that

$$\sum_{k=1}^{\infty} a_{nk} = 1 \qquad (n = 1, 2, \cdots). \tag{5.28}$$

Hence A is not only multiplicative but is also regular. To show that A satisfies (4.152), put, for each $n = 1, 2, 3, \cdots$,

$$F_n(\theta) = \sum_{k=1}^{\infty} |\, \mathcal{R}\,(e^{i\theta} a_{nk}) \,| \qquad -\pi < \theta \leqq \pi. \tag{5.31}$$

If we substitute in (5.31) the values of a_{nk} as given by (5.21), it becomes apparent that $F_n(\theta)$ has period $\pi/2n$. Moreover if $-\pi/2n < \theta \leqq \pi/2n$, then $\mathcal{R}\,(e^{i\theta} a_{nk}) \geqq 0$ so that

$$F_n(\theta) = \sum_{k=1}^{\infty} \mathcal{R}\,(e^{i\theta} a_{nk}) = \mathcal{R}\,(e^{i\theta} \sum_{k=1}^{\infty} a_{nk}) = \mathcal{R}\,(e^{i\theta}) = \cos\theta. \tag{5.32}$$

Therefore $F_n(\theta) \leqq 1$ when $-\pi/2n < \theta \leqq \pi/2n$, and periodicity gives $F_n(\theta) \leqq 1$ for all θ. This implies (4.152). Thus we have a regular matrix A having the property that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n and also the property $\lim \sum_{k=1}^{\infty} |\, a_{nk} \,| = \pi/2$. The matrix A of (5.21) is specially constructed to have these properties, and further properties of A are irrelevant to our discussion. We remark, however, that the geometric series $1 + z + z^2 + \cdots$ is summable A to the "correct" value $1/(1-z)$ for each $z \neq 1$ which lies within or on the boundary of the unit circle $|\, z \,| = 1$, and hence that A has at least some of the attributes of an orthodox method of summability.

6. Necessary conditions. We now prove

THEOREM 6.1. *If A is a multiplicative matrix with the property*

$$\Omega(y) \leqq \Omega(x) \qquad x_n \text{ real and bounded} \tag{6.11}$$

then

$$\lim_{n\to\infty} |\, \sum_{k=1}^{\infty} a_{nk} \,| = |\, \rho \,| \leqq 1 \tag{6.12}$$

and

$$\limsup_{n\to\infty} \sum_{k=1}^{\infty} |\, a_{nk} \,| \leqq \pi/2. \tag{6.13}$$

Our hypothesis and Theorem 4.1 give (4.14) and (6.12) follows.

To establish (6.13), we note that our hypothesis and Theorem 4.1 yield (4.152). Let $\epsilon > 0$. Choose N so great that

$$\text{(6.14)} \qquad \text{l. u. b.}_{-\pi < \theta \leqq \pi} \sum_{k=1}^{\infty} | \mathcal{R}(e^{i\theta} a_{nk}) | \leqq 1 + \epsilon \qquad n \geqq N.$$

Then for each $n \geqq N$, we have

$$\text{(6.15)} \qquad \sum_{k=1}^{\infty} | \mathcal{R}(e^{i\theta} a_{nk}) | \leqq 1 + \epsilon \qquad -\pi < \theta \leqq \pi.$$

Since $\sum_{k=1}^{\infty} | a_{nk} | < \infty$, the series in the left member of (6.15) converges uniformly. Hence we can integrate termwise to obtain

$$\text{(6.16)} \qquad \sum_{k=1}^{\infty} \int_{-\pi}^{\pi} | \mathcal{R}(e^{i\theta} a_{nk}) | \, d\theta \leqq 2\pi(1 + \epsilon).$$

If we put $a_{nk} = | a_{nk} | \, e^{i\theta_{nk}}$, we obtain

$$\int_{-\pi}^{\pi} | \mathcal{R}(e^{i\theta} a_{nk}) | \, d\theta = | a_{nk} | \int_{-\pi}^{\pi} | \cos(\theta_{nk} + \theta) | \, d\theta.$$

Since $\cos \phi$ has period 2π,

$$\int_{-\pi}^{\pi} | \cos(\theta_{nk} + \theta) | \, d\theta = \int_{-\pi}^{\pi} | \cos \theta | \, d\theta = 4.$$

Therefore

$$\text{(6.17)} \qquad \sum_{k=1}^{\infty} | a_{nk} | \leqq (\pi/2)(1 + \epsilon) \qquad n \geqq N$$

and (6.13) follows.

The condition (6.13) is of course not sufficient to ensure (6.11). The matrix A of (5.21) shows, however, that Theorem 6.1 will not hold if the constant $\pi/2$ in the right member of (6.13) is replaced by a lesser constant.

7. Norms of sequences and of their transforms. Let the *norm* of a sequence s_n of complex numbers be denoted by

$$\text{(7.01)} \qquad L(s) = \limsup_{n \to \infty} | s_n |.$$

It is easy to show that a matrix A is such that $L(y) \leqq L(x)$ for each bounded complex sequence x_n if and only if the two conditions

$$\text{(7.02)} \qquad \lim_{n \to \infty} a_{nk} = 0 \qquad (k = 1, 2, \cdots),$$

$$\text{(7.03)} \qquad \limsup_{n \to \infty} \sum_{k=1}^{\infty} | a_{nk} | \leqq 1$$

are satisfied. Characterization of matrices A such that $L(y) \leqq L(x)$ for each

bounded real divergent sequence x_n is not so simple. Methods of proof used in §§ 2-3 suffice to prove the following three lemmas.

LEMMA 7.1. *In order that* $L(y) \leqq L(x)$ *wherever* x_n *is real, bounded, and divergent, it is necessary and sufficient that* $L(y) \leqq L(x)$ *whenever* x_n *is real and bounded.*

LEMMA 7.2. *In order that* $L(y) \leqq L(x)$ *whenever* x_n *is real and bounded, it is necessary that* (2.21) *and* (2.22) *hold.*

LEMMA 7.3. *In order that* $L(y) \leqq L(x)$ *whenever* x_n *is real and bounded, it is necessary and sufficient that* (2.21) *and* (2.22) *hold and* $L(y) \leqq 1$ *whenever* $-1 \leqq x_n \leqq 1$.

Using these lemmas and lemmas of § 3, we obtain the following theorem.

THEOREM 7.4. *In order that a matrix* A *may be such that* $L(y) \leqq L(x)$ *for each real bounded sequence* x_n, *it is necessary and sufficient that* (2.21) *and* (2.22) *hold, and that one of* (4.13), (4.14), (4.142), (4.15), *and* (4.152) *holds.*

8. Sequence-to-function transformations. Let E be a subset of a space in which the notion of limit of a sequence is defined, and let t_0 be a limit point of E not belonging to E. For each $k = 1, 2, \cdots$, let $a_k(t)$ be a function defined for $t \in E$. The transformation

$$\mathcal{A} \qquad y(t) = \sum_{k=1}^{\infty} a_k(t) x_k$$

associates a transform $y(t)$ with each sequence x_n for which the series converges when $t \in E$. The transform $y(t)$ is convergent if $\lim_{t \to t_0} y(t)$ exists, it being always understood that $t \to t_0$ over the set E. The oscillation of $y(t)$ is defined by

$$\Omega(y) = \limsup_{t, u \to t_0} | y(t) - y(u) |. \tag{8.1}$$

We leave it to the reader to see that each definition and theorem of this paper involving A can be changed into one involving $\mathcal{A}$ simply by replacing n by t so that y_n becomes $y(t)$, a_{nk} becomes $a_k(t)$, and $n \to \infty$ becomes $t \to t_0$. For example, the condition

$$\limsup_{t \to t_0} \text{ l. u. b.}_{U_k(t) = \pm 1} \left| \sum_{k=1}^{\infty} a_k(t) U_k(t) \right| \leqq 1 \tag{8.2}$$

is a necessary and sufficient condition that a multiplicative transformation $\mathcal{A}$ be such that $\Omega(y) \leqq \Omega(x)$ for each real bounded sequence x_n.

CORNELL UNIVERSITY,
ITHACA, NEW YORK.

GROUPS WITH ABELIAN NORM QUOTIENT GROUP.*[1]

By REINHOLD BAER.

The norm of a group G consists of all those elements of G which transform every subgroup of G into itself. It is the object of this note to describe the group of automorphisms which are induced in the norm of G by the elements of G, provided the norm quotient group is abelian.

1. In this section we are going to enumerate certain concepts and results which have been derived elsewhere and which will be needed in the future.

If G is a group, then its *central* $C(G)$ consists of all those elements u which satisfy: $uxu^{-1} = x$ for every x in G; and similarly the *norm* $N(G)$ of G consists of all those elements u which satisfy: $uXu^{-1} = X$ for every subgroup X of G.

If $C(G) \neq N(G)$, and if the norm quotient group is abelian, then G is the direct product of its primary components [2] and $N(G)$ is the direct product of the norms of the primary components.[3] For this reason we may assume that G is a p-group.

If G is a p-group, then the order of the element g in G is a power $p^{n(g)}$ of the prime number p. If N is a normal (or self conjugate or invariant) subgroup of G, then every element g in G induces an automorphism in N. The order of this automorphism is a power $p^{a(g)}$ of the prime number p and we put $m(g) = n(g) - a(g)$. Both the numbers $a(g)$ and $m(g)$—and several other functions which will be introduced shortly—depend on N. But the omission of the variable N will not lead to any confusion.—It should be noted that the elements g with $a(g) = 0$ form the centralizer $C(N < G)$ of N in G. Consequently $a(g)$ is the smallest integer i so that g^{p^i} is an element in $C(N < G)$ and this property may be used as a definition of $a(g)$ even, if N is not a normal subgroup of G.

The following is a convenient restatement of a fundamental theorem on groups with cyclic norm quotient group.[4, 5]

* Received September 6, 1938.

[1] Presented to the American Mathematical Society, November 25-26, 1938.

[2] Cp. Satz 1, § 1 of R. Baer, "Gruppen mit vom Zentrum wesentlich verschiedenem Kern und abelscher Faktorgruppe nach dem Kern," *Comp. Math.*, vol. 4 (1936), pp. 1-77; in the following referred to as BWK.

[3] Cp. Satz 4, § 3 of R. Baer, "Der Kern, eine characteristische Untergruppe," *Comp. Math.*, vol. 1 (1934), pp. 254-283; in the following referred to as BK.

[4] Cp. Satz 7, § 5 of BK.

[5] The following two notations will be used throughout:
$\{\cdots\}$ is the subgroup, generated by the enclosed elements and element sets;
$[u, v] = uvu^{-1}v^{-1}$.

(1.1) *The element g and the subgroup N of the p-group G satisfy:*

(a) *N is a normal subgroup of $\{N, g\}$;*

(b) *$N \leq N(\{N, g\}) < \{N, g\}$;*

if, and only if,

(i) *N is abelian;*

(ii) *$n(x) \leq m(g)$ for every x in N;*

(iii) *$c(g) = g^{p^{m(g)}}$ is an element in N* (so that *$a(g) \leq m(g)$ by* (ii));

(iv) *$g^{-1}xg = xc(g)^{h(x,g)}$ for every x in N;*

(v) *$g^{-1}eg = ec(g)$ for some e in N;*

(vi) *$1 < m(g)$ in case $p = 2$.*

It should be noted that (ii) implies that the orders of the elements in N are bounded; and we are going to denote the *maximum order of the elements in N* by p^m. Then (iii) implies that $a(g) \leqq m$.

Another consequence of the theorem on groups with cyclic norm quotient group is the following statement:

(1.2) *If the element g and the subgroup N of the p-group G satisfy the conditions* (a) *and* (b) *of* (1.1), *then*

$$N(\{N, g\}) = \{N, g^{p^{a(g)}}\}.$$

Note that the definition of $c(g)$, as given in (iii) of (1.1), is valid for any element g in G. Then the three statements: $c(g) = 1$, $a(g) = 0$ and "g is in $C(N < G)$" are equivalent.

The proofs of the following formulae may be omitted, as they are practically identical with the proofs of some slightly more special statements, given elsewhere.[6]

(1.3) *If the element g and the subgroup N of the p-group G satisfy the conditions* (a) *and* (b) *of* (1.1), *and if x is an element in N, then*

$$(xg)^i = x^i c(g)^{\frac{-i(i-1)}{2} h(x,g)} g^i.$$

(1.4) *Suppose that the subgroup N and both the elements u and v of the p-group G satisfy the conditions* (a) *and* (b) *of* (1.1), *and that $\{N, u, v\}/N$ is abelian. Then*

(1) $$[v^i, u] = [v, u]^i c(v)^{\frac{i(i-1)}{2} h([u,v],v)}.$$

(2) $$(uv)^i = [v, u]^{\frac{i(i-1)}{2}} c(v)^{\frac{i(i-1)(i-2)}{6} h([u,v],v)} \times c(u)^{\frac{(2i-1)i(i-1)}{6} h([u,v,]u) + \frac{i(i-1)(i-2)(3i-1)}{24} h([v,u],v)h(c(v),u)} u^i v^i.$$

[6] (4), (5) and (6) of § 2 of BWK.

$$(3)\quad c(u)^{h(c(v),u)} = 1 \text{ for } p \neq 2;$$
$$= c(v)^{h([u,v],v)2^{m(v)-1}} \text{ for } p = 2.$$

In order to prove the last of these formulae one has to consider that either $c(v) = 1$ or $a(v) \leq n(Nv) \leq m(v) = n(v) - a(v)$ and $n(x) \leq m(v)$ for every x in N. [(3) is actually a special case of (1).]

Applying formula (3) on (2) one gets the following formula:

$$(2')\quad (uv)^i = [v,u]^{\frac{i(i-1)}{2}} c(v)^{\frac{i(i-1)(i-2)}{6} h([u,v],v)} \times c(u)^{\frac{(2i-1)i(i-1)}{6} h([u,v],u)} u^i v^i \text{ for } p \neq 2$$

$$(uv)^i = [v,u]^{\frac{i(i-1)}{2}} c(u)^{\frac{(2i-1)i(i-1)}{6} h([u,v],u)} \times c(v)^{\frac{i(i-1)(i-2)}{6} h([u,v],v) + \frac{i(i-1)(i-2)(3i-1)}{24} 2^{m(v)-1} h([u,v],v)^2} u^i v^i \text{ for } p =$$

2. If the normal subgroup N of the p-group G is contained in the norm $N(G)$ of G, and if G/N is abelian, then there are two classes of elements in G on which the results of section **1** cannot be applied, namely firstly the elements in $C(N < G)$ and secondly the elements g so that $\{N, g\}$ is a hamiltonian group. This latter possibility can only occur, if $p = 2$, and shall be excluded from our investigation.[7] If we exclude furthermore that $G = C(N < G)$ or equivalent: that $N \leq C(G)$, then there exist certainly elements in G which satisfy together with N the conditions (a) and (b) of (1.1) so that in particular the orders of the elements in N are bounded. Denoting by p^m the maximum order of the elements in N, we introduce a new function:

$$d(g) = g^{p^m} \text{ for } g \text{ in } G.$$

LEMMA 2.1. *Suppose that the normal subgroup N of the p-group G is contained in the norm, but not in the central of G, that G/N is abelian, and that G does not contain elements w so that $\{N, w\}$ is hamiltonian.*

(1) *If g is in G and x in N, then*

$$d(xg) = d(g) \text{ for } p \neq 2;$$
$$= c(g)^{2^{m-1}h(x,g)} d(g) \text{ for } p = 2.$$

$$(2)\quad [d(v), u] = 1 \text{ for } p \neq 2;$$
$$= c(v)^{2^{m-1}h([u,v],v)} \text{ for } p = 2.$$

[7] A paper on "Almost Hamiltonian Groups," which is going to appear in an early issue of *Compositio Mathematica*, will be devoted to the study of this class of groups.

$$
\begin{aligned}
(3)\quad d(uv) &= d(u)d(v) \text{ for } p \neq 2,\ p \neq 3;\\
&= c(u)^{3^{m-1}h([v,u],u)}c(v)^{3^{m-1}h([u,v],v)}d(u)d(v) \text{ for } p = 3;\\
&= [u,v]^{2^{m-1}}c(u)^{2^{m-1}h([v,u],u)}d(u)d(v) \text{ for } p = 2 < m;\\
&= [u,v]^{2}c(u)^{2h([v,u],u)}c(v)^{2^{m(v)-1}h([u,v],v)^2}d(u)d(v) \text{ for } p = 2 = m;\\
&= [u,v]c(u)^{h([v,u],u)}d(u)d(v) \text{ for } p = 2,\ m = 1.^{8}
\end{aligned}
$$

These formulae are consequences of the fact that p^m is the maximum order of the elements in N, that $c(g)$ and $[u, v]$ are elements in N, if one puts $i = p^m$ in the corresponding formulae of section **1**.

If one puts $d'(g) = g^{p^{m+1}}$, then one may prove exactly as the lemma has been proved:

COROLLARY 2.2. *If the assumptions of the lemma are satisfied, then $d'(g)$ belongs to $C(G)$; $d'(xg) = d'(g)$ for x in N and g in G; $d'(uv) = d'(u)d'(v)$ for u, v in G.*

THEOREM 2.3. *If G is a p-group with abelian norm quotient group, and if norm and central of G are different, then the orders of the elements in G are bounded.*

Proof. If $N(G)$ is contained in a hamiltonian group whose quotient group mod $N(G)$ is cyclic, then it has been proved elsewhere [9] that the orders of the elements in G do not exceed 8. If G does not contain a hamiltonian subgroup with these properties, then it follows from Corollary 2.2 that $g^{p^{m+1}}$ is an element in $C(G)$ for every g in G. But as $C(G) \leqq N(G)$ and as p^m is the maximum order of the elements in $N(G)$, it follows that $g^{p^{2m+1}} = 1$ for every g in G.

3. A group G shall be called *i-homomorph*, if $(xy)^i = x^i y^i$ for any two elements x and y in G, i. e. if mapping upon the i-th powers defines a homomorphism of G upon a subgroup of G. Let us denote furthermore by $(G; x^i = 1)$ the set of elements x in G whose order is a divisor of the integer i. If the group G is i-homomorph, then $(G; x^i = 1)$ is a characteristic subgroup of G and an isomorphism of $G/(G; x^i = 1)$ upon a subgroup of G is defined in mapping every element of G upon its i-th power.

LEMMA 3.1. *Suppose that the p-group G and its subgroup N satisfy the assumptions of Lemma 2.1, that p^m is the maximum order of the elements in N, and that G is p^m-homomorph.*

---

[8] The statement in (7), § 2 of BWK is not correct. But this affects only some of the considerations of § 10 of BWK.

[9] Cp. footnote 7.

(a) *If* $p = 2$, *then* $a(g) < m$ *for every* g *in* G, *and* $[u, v]^{2^{m-1}} = 1$ *for* u, v *in* G.
(b) *The elements* $d(g) = g^{p^m}$ *for* g *in* G *form a subgroup* $D(N < G)$ *of the central* $C(G)$ *of* G.
(c) $N \leqq (G; x^{p^m} = 1) \leq C(N < G)$.
(d) *If* u, v *and* w *are independent* mod $(G; x^{p^m} = 1)$, *then* $[u, v]w = w[u, v]$.

Proof. Suppose first that $p = 2$. Then it follows from Lemma 2.1, (1) that $c(g)^{2^{m-1}h(x,g)} = 1$ for every g in G and every x in N, since $d(x) = 1$, $d(xg) = d(x)d(g) = d(g)$. Since by (1.1) there exist elements x in N so that $h(x, g) = 1$—at least if $c(g) \neq 1$—this implies that $c(g)^{2^{m-1}} = 1$ for every g in G. As $n(c(g)) = a(g)$, we have therefore proved that $a(g) < m$. Furthermore it follows from (1.1), (vi), that $2 \leq m(g)$, if $0 < a(g)$. Now it is a consequence of Lemma 2.1, (3) that

$$d(u)d(v) = d(uv) = [u, v]^{2^{m-1}}d(u)d(v)$$

and this completes the proof of (a).

It follows from Lemma 2.1, (2) and from the statement (a)—proved just now—that $[d(v), u] = 1$ for every u and v in G and this proves the assertion (b).

Since $m \leq m(g)$ for every g in G with $c(g) \neq 1$, it follows that $c(g) = d(g)^{p^{m(g)-m}}$ for these g in G. Consequently $c(g) = 1$, if $d(g) = 1$ and this proves (c).

If u, v and w are independent mod $(G; x^{p^m} = 1)$, then $d(u)$, $d(v)$ and $d(w)$ are independent elements in the abelian group $C(G)$. Since G is an extension of the abelian group N by the abelian group G/N, there holds the identity:[10]

$$\begin{aligned} 1 &= [u, v]w[v, u]w^{-1}[v, w]u[w, v]u^{-1}[w, u]v[u, w]u^{-1} \\ &= c(w)^{h([u,v],w)}c(u)^{h([v,w],u)}c(v)^{h([w,u],v)} \text{ by (1.1).} \end{aligned}$$

But since the d's are independent, it follows that the c's are independent too, i. e. $1 = c(w)^{h([u,v],w)}$ and this proves (d).

COROLLARY 3.2. *Suppose that the* p*-group* G *and its subgroup* N *satisfy the conditions of Lemma* 2.1, *and that* p^m *is the maximum order of the elements in* N.

(a) *If* $p \neq 2$ *and* $p \neq 3$, *then* G *is* p^m*-homomorph.*
(b) *If* $p = 3$ *and* G *is* 3^m*-homomorph, then* $[u, v]^{3^{m-1}}$ *is in* $C(G)$ *for any pair of* mod $(G; v^{3^m} = 1)$ *independent elements* u *and* v.

[10] For a verification of this identity cp. e. g. R. Baer, "Erweiterung von Gruppen und ihren Isomorphismen," *Mathematische Zeitschrift*, vol. 38 (1934), pp. 375-416; in particular p. 407.

(c) *If $p = 2$, then G is 2^m-homomorph if, and only if, $a(g) < m$ for every g in G and $[u, v]^{2^{m-1}} = 1$ for every u, v in G.*

These are consequences of Lemma 3.1 and Lemma 2.1.

4. If G and its subgroup N satisfy the assumptions of Lemma 3.1, then $(G; x^{p^m} = 1)$ is contained in the centralizer of N in G. Since under these assumptions the orders of the elements in G are bounded, it follows that the orders of the elements in $C(N < G)$ are bounded, and that the maximum order of the elements in $C(N < G)$ has the form p^{m+q} where q is a not-negative integer.

THEOREM 4.1. *Suppose that the p-group G and its normal subgroup N satisfy the assumptions of Lemma 3.1, that p^m is the maximum order of the elements in N, and that p^{m+q} is the maximum order of the elements in $C(N < G)$.*

(i) *$c(g) = g^{p^{m+q}}$ for every g in G.*

(ii) *$C(N < G) = (G; x^{p^{m+q}} = 1)$.*

(iii) *An isomorphism of the group of automorphisms of N, which the elements of G induce in N, upon a subgroup $K(N < G)$ of the cross cut of N and $C(G)$ is defined in mapping the automorphism of N, induced by the element g of G, upon the p^{m+q}-th power $c(g)$ of g.*

(iv) *N is mod the cross cut of N and $C(G)$ a cyclic group whose order equals the maximum order of the automorphisms of N which are induced by elements of G.*

(v) *To every element x of N there exists an integer $h(x)$ so that $g^{-1}xg = xc(g)^{h(x)}$ for every g in G.*

Proof. If g is an element in G, but not in $(G; x^{p^m} = 1)$, then either $0 < a(g)$ and $m \leq m(g)$ by (1.1), (ii) or $0 = a(g)$ and $m < n(g) = m(g)$. In either of these cases we have therefore $m \leq m(g)$ and $c(g) = d(g)^{p^{m(g)-m}}$.

If G is a cyclic group mod $(G; x^{p^m} = 1)$, then our Theorem is a consequence of (1.1) and of Lemma 3.1.

Suppose now that the elements u and v in G are independent mod $(G; x^{p^m} = 1)$ so that neither of these elements is contained in $(G; x^{p^m} = 1)$ whereas the cross cut of the cyclic groups, generated by $(G; x^{p^m} = 1)u$ and $(G; x^{p^m} = 1)v$ respectively, consists of the group-unit only. Then it follows from section **3**, that the cross cut of the cyclic groups, generated by $d(u)$ and

$d(v)$ respectively, consists of the group-unit only, whereas neither $d(u)$ nor $d(v)$ equals 1. If now x is any element in N, then

$$\begin{aligned}(uv)^{-1}x(uv) &= xc(uv)^{h(x,uv)} \\ &= v^{-1}xc(u)^{h(x,u)}v = xc(u)^{h(x,u)}c(v)^{h(x,v)},\end{aligned}$$

since $c(u)$ is by Lemma 3.1 an element in the central of G. Hence we find:

$$\begin{aligned}d(u)^{p^{m(uv)-m}h(x,uv)}d(v)^{p^{m(uv)-m}h(x,uv)} &= d(uv)^{p^{m(uv)-m}h(x,uv)} \\ = c(uv)^{h(x,uv)} &= c(u)^{h(x,u)}c(v)^{h(x,v)} \\ &= d(u)^{p^{m(u)-m}h(x,u)}d(v)^{p^{m(v)-m}h(x,v)},\end{aligned}$$

since both $d(u)$ and $d(v)$ are in $C(G)$. Since u and v are assumed to be independent mod $(G; x^{p^m} = 1)$, it follows now that

$$\begin{aligned}d(u)^{p^{m(uv)-m}h(x,uv)} &= d(u)^{p^{m(u)-m}h(x,u)}, \\ d(v)^{p^{m(uv)-m}h(x,uv)} &= d(v)^{p^{m(v)-m}h(x,v)}\end{aligned}$$

or

$$\begin{aligned}h(x,uv)p^{m(uv)-m} &\equiv h(x,u)p^{m(u)-m} \bmod p^{n(u)-m}, \\ &\equiv h(x,v)p^{m(v)-m} \bmod p^{n(v)-m}.\end{aligned}$$

Assuming finally—as we may do without loss of generality—that $n(u) \leq n(v)$, we find that

(4.1.1) $\qquad h(x,v)\,p^{m(v)-m} \equiv h(x,u)\,p^{m(u)-m} \bmod p^{n(u)-m}$ for x in N.

If v is an element in $C(N < G)$, then $c(v) = 1$, $a(v) = 0$ and $h(x,u)\,p^{m(u)-m} \equiv 0 \bmod p^{n(u)-m}$. Since there exist by (1.1) elements x with $h(x,u) = 1$—at least if $0 < a(u)$—this implies that

$$n(u) - m \leq m(u) - m = n(u) - a(u) - m \text{ or } a(u) = 0.$$

Thus we have proved

(4.1.2) if u and v are independent mod $(G; x^{p^m} = 1)$, if $n(u) \leq n(v)$, and if v belongs to $G(N < G)$, then u belongs to $C(N < G)$.

Since $N \leq (G; x^{p^m} = 1) \leq C(N < G)$ by Lemma 3.1, (c), and since G/N is abelian and the orders of its elements are bounded, it follows that either $C(N < G) = (G; x^{p^m} = 1)$ or that $C(N < G)/(G; x^{p^m} = 1)$ as well as $(G; x^{p^{m+q}} = 1)/(G; x^{p^m} = 1)$ is generated by a set of independent elements,[11] and thus (ii) is a consequence of the above statement (4.1.2).

If g is an element in $C(N < G)$, then (i) is a consequence of (ii). If g is not contained in $C(N < G)$, then $a(g)$ is the smallest positive number so

[11] Cp. e.g. BK, Lemma, pp. 274-275.

that $g^{p^{a(g)}}$ is an element in $C(N < G)$, and by (ii) it follows therefore that $a(g)$ is the smallest positive number so that $n(g) - a(g) \leq m + q$. Hence $m(g) = m + q$ and this completes the proof of (i).

If u and v are any two elements in G, then

$$c(uv) = (uv)^{p^{m+q}} = d(uv)^{p^q} = d(u)^{p^q}\, d(v)^{p^q} = c(u)\, c(v)$$

as follows from (i) and the fact that all the elements $d(g)$ are in $C(G)$. Since any two elements in G induce in N the same automorphism if, and only if, they are in the same coset mod $C(N < G)$, (iii) follows now from (ii).

From (i) and (4. 1. 1) it follows that

$$h(x, v)\, p^q \equiv h(x, u)\, p^q \bmod p^{n(u)-m},$$

if x is an element in N, u and v are independent mod $(G; x^{p^m} = 1)$, $n(u) \leq n(v)$. If $0 < a(u)$, then this involves

$$h(x, v) \equiv h(x, u) \bmod p^{a(u)}.$$

If we choose in particular u as an element of maximum order in G, then it follows from (i) to (iii) that u is an element of maximum order mod $C(N < G)$. Since there exists a basis of $G/C(N < G)$, the above congruence implies:

If the set B is a basis of G mod $C(N < G)$, then there exists to every element x in N an integer $h(x)$ so that

$$g^{-1}xg = x\, c(g)^{h(x)} \text{ for every } g \text{ in } B.$$

Since $g^{-i}xg^i = xc(g)^{ih(x)} = xc(g^i)^{h(x)}$, since B generates G mod $C(N < G)$, and since every $c(g)$ is an element of the central of G, this implies (v).

Denote now by M the cross cut of N and $C(G)$ and by M^* the set of all those elements in N which permute with the element w of maximum order in G. Clearly $M \leq M^*$. If conversely x is an element in M^*, then $x = w^{-1}xw = xc(w)^{h(x)}$, or $h(x) \equiv 0 \bmod p^{a(w)}$. But since w is an element of maximum order in G, it follows from (ii) that w is an element of maximum order mod $C(N < G)$, i. e. an element of greatest $a(w)$ and this proves that $c(g)^{h(x)} = 1$ for every g. Hence x is an element in M or $M = M^*$. If finally e is an element in N so that $h(e, w) = 1$—such an element exists by (1. 1)—then for every x in N the element $xe^{-h(x,\, w)}$ is contained in $M^* = M$. Hence N is generated mod $M^* = M$ by the element e and this proves (iv).

The following statement has been derived during the proof of the Theorem; and it is a fairly obvious consequence of (i) too.

COROLLARY 4.2: *If the p-group G and its normal subgroup N satisfy the assumptions of Lemma 3.1, and if the element g of G is not contained in $C(N<G)$, then $g^{p^{a(g)}}$ is an element of maximum order in $C(N<G)$.*

The most important special case which is contained in Corollary 3.2 and Theorem 4.1 may be stated as follows:

If G is a p-group with abelian norm quotient group, if norm and central of G are different, if p^r is the maximum order of the elements in $C(N(G)<G)$, and if $p\neq 2$ and $p\neq 3$, then

p^r is finite (though G may be an infinite group);

$N(G)/C(G)$ is a cyclic group whose order is the maximum order in the group $A(G)$ of automorphisms, induced in $N(G)$ by elements in G;

the correspondence: $g\rightarrow g^{p^r}$ defines an isomorphism of $A(G)$ upon a subgroup of $C(G)$;

$g^{-1}xg=x\,(g^{p^r})^{h(x)}$ for every x in $N(G)$, g in G where $h(x)$ does not depend on g;

$C(N(G)<G)$ consists exactly of the elements x in G which satisfy: $x^{p^r}=1$.

That the assumptions $p\neq 2$ and $p\neq 3$ are essential, may be seen from suitable examples.[12]

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

[12] Cp. §§ 9, 10 of BWK.

CONDITION THAT A FINITE GROUP BE MULTIPLY ISOMORPHIC WITH EACH OF ITS IRREDUCIBLE REPRESENTATIONS.* [1]

By Louis Weisner.

A finite group may be multiply isomorphic with each of its irreducible representations as a linear group. A necessary condition that a group be simply isomorphic with at least one of its irreducible representations is that the central of the group be cyclic, but this condition is not sufficient.[2] In fact, a group may have no simply isomorphic irreducible representation even if its central is the identity group, as Burnside showed by an example.[3] Burnside also showed that a group has at least one simply isomorphic irreducible representation if it does not contain two distinct minimal invariant subgroups whose orders are powers of the same prime, but he recognized that this condition is not necessary.

The problem of determining a necessary and sufficient condition that a finite group be multiply isomorphic with each of its irreducible representations has been treated at length by Shoda [4] whose solution, it appears to me, was unnecessarily complicated by the introduction of extraneous concepts. Indeed, Shoda's final criterion is hardly in a form in which it may be conveniently applied to a given abstract group.

I propose, in this paper, to give a relatively simple solution of the problem by establishing the following theorem.

Theorem. *A necessary and sufficient condition that a finite group G be multiply isomorphic with each of its irreducible representations as a linear*

* Received by the editors September 1, 1938.

[1] Presented to the American Mathematical Society, October 29, 1938.

[2] That the condition is sufficient for a prime-power group was proved by W. B. Fite, "Groups whose orders are powers of a prime," *Transactions of the American Mathematical Society*, vol. 7 (1906), p. 67.

[3] W. Burnside, *Theory of Groups of Finite Order*, second edition, 1911, Note F, p. 476.

[4] K. Shoda, "Ueber direkt zerlegbare Gruppen," *Journal of the Faculty of Science*, Imperial University of Tokyo, Section 1, vol. 2 (1929), pp. 51-72; "Bemerkungen über vollständige reduzible Gruppen, *ibid.*, pp. 203-209.

group is that G contain a subgroup P whose order is a power of a prime p, such that (1) P is generated by the minimal invariant subgroups of G whose orders are powers of p; and (2) every subgroup of index p of P contains an invariant subgroup of G different from the identity group.

We shall confine our attention to composite groups since a simple group G has no subgroup P with the stated properties, and a simple group always has a simply isomorphic irreducible representation.

A composite group G contains one or more minimal invariant subgroups each of which is a direct product of simply isomorphic simple groups. Supposing that $p_1, \cdots, p_m$ are the distinct prime factors of the order of G, let P_i be the group generated by the minimal invariant subgroups of G whose orders are powers of p_i; in the event that no such minimal invariant subgroup exists, let $P_i = 1$. Since two minimal invariant subgroups of a group have no element in common besides the identity, each element of one is commutative with every element of the other. Therefore P_i is either the identity group or an abelian group of type $(1, 1, 1, \cdots)$. Let S be the group generated by those minimal invariant subgroups of G which are direct products of simply isomorphic simple groups of composite order; in the event that there are no such subgroups, let $S = 1$. The group generated by all the minimal invariant subgroups of G is

$$D = S \times P_1 \times \cdots \times P_m. \tag{1}$$

It is to be observed that S is either the identity group or the direct product of simple groups of composite order; while P_i is either the identity group or the direct product of groups of order p_i. Since G is composite at least one of the factors of D is different from the identity group.

If $\mathbf{\Sigma}$ is any irreducible representation of S, and $\mathbf{\Pi}_i$ is any irreducible representation of P_i, then

$$\Delta = \mathbf{\Sigma} \times \mathbf{\Pi}_1 \times \cdots \times \mathbf{\Pi}_m \tag{2}$$

is an irreducible representation of D; the direct product of two or more linear groups which operate on distinct symbols being obtained by forming all possible products (Kronecker products) of the substitutions of the factors.[5]

Now suppose that each P_i which is different from the identity group has a subgroup L_i, of index p_i in P_i, that contains no invariant subgroup of G besides the identity group. Since P_i is an abelian group of type $(1, 1, 1, \cdots)$, each of its irreducible representations operates on a single variable; and there

[5] Burnside, *loc. cit.*, p. 478; Shoda, *loc. cit.*, p. 64.

is an irreducible representation of P_i in which all the elements of any selected subgroup of index p_i, and no others, are represented by the identity substitution. For each $P_i \neq 1$, let Π_i be an irreducible representation of P_i in which the elements of L_i are the only ones that are represented by the identity substitution; for each $P_i = 1$, let Π_i be the identical representation. Since S, in (1), is either the identity group or the direct product of simple groups of composite order, S has an irreducible representation Σ simply isomorphic with S. From these representations we form the representation Δ of D defined by (2).

There is at least one irreducible representation of G in which the subgroup corresponding to D, when completely reduced, has Δ as a component. We proceed to prove that if Γ is such a representation of G, Γ is simply isomorphic with G. If Γ is multiply isomorphic with G, G has an invariant subgroup $H \neq 1$ whose elements are represented in Γ by the identity substitution. H has a subgroup J which is a minimal invariant subgroup of G. J cannot be the direct product of simple groups of composite order; for in that case J would be a subgroup of S which is represented in Γ by a group simply isomorphic with S. Hence J is a prime-power group; and J is a subgroup of one of the P's, say P_i. When the subgroup of Γ corresponding to D is completely reduced, the component Δ occurs. In Δ each element of L_i, and no other element of P_i, is represented by the identity substitution. Therefore L_i contains J, contrary to the assumption that no invariant subgroup of G, except the identity group, is contained in L_i. Γ is therefore simply isomorphic with G. It follows that if G is multiply isomorphic with each of its irreducible representations, at least one of the P's must have the properties stated in the theorem.

To prove the converse, suppose that P, of order p^α, is one of the P's that occurs in (1), and that every subgroup of index p of P contains an invariant subgroup of G different from the identity; hence $\alpha \geq 2$. We shall prove that no irreducible representation of G is simply isomorphic with G. Let Γ be an irreducible representation of G, and let Π be the subgroup of Γ that corresponds to P. We may suppose that Π is simply isomorphic with P since, in the contrary case, Γ would not be simply isomorphic with G. Let Π' be one of the irreducible representations of P that arises when Π is completely reduced. If Π' is the identical representation of P, every element of P is represented by the identity substitution in Π'. If Π' is not the identical representation of P, exactly $p^{\alpha-1}$ elements of P are represented in Π' by the identity substitution. In any case there is at least one subgroup L of order $p^{\alpha-1}$ of P whose elements are represented in Π' by the identity substitution. If Λ is the subgroup of Γ that corresponds to L, Λ has a linear invariant since the identical representa-

tion arises when Λ is completely reduced. By hypothesis some subgroup $F \neq 1$ of L is invariant under G. If Φ is the corresponding subgroup of Γ, Φ has a linear invariant since Φ is a subgroup of Λ. Now an invariant subgroup of an irreducible linear group cannot have a linear invariant unless all its substitutions are the identity substitution.[6] Therefore each of the substitutions of Φ is the identity substitution, and Γ is multiply isomorphic with G.

HUNTER COLLEGE OF THE CITY OF NEW YORK.

[6] If B is an invariant subgroup of an irreducible linear group A which operates on n variables, the substitutions of A transform each linear invariant of B into a linear invariant of B. Therefore a set of r linear invariants of B in terms of which all the linear invariants of B may expressed linearly forms a reduced set for A. Since A is irreducible, $n = r$, and each of the variables operated on by A is an invariant of B.

A different proof is given by A. Koulakoff, "Sur un théorème de la théorie des caractères," (Russian with French summary), *Recueil math. Moscou*, vol. 36 (1929), pp. 129-134.

ADDITIVE ARITHMETICAL FUNCTIONS AND STATISTICAL INDEPENDENCE.*

By PAUL ERDÖS and AUREL WINTNER.

Introduction. The object of this paper is to solve the problem[1] as to the existence of an asymptotic distribution function of an arbitrary (real) additive arithmetical function $f(n)$, and to embed this problem into the general theory of infinite convolutions.[2] This aim will be reached by an approach which subjects $f(n)$ to no a priori limitation whatever. Actually, the sufficient condition found *loc. cit.*[1], III for the existence of the asymptotic distribution function turns out to be a necessary condition as well.

The result to be obtained is to the effect that the statistical independence of the components which belong to the different prime numbers is such as to make the condition of convergence in relative measure not only sufficient but also necessary for the existence of an asymptotic distribution function.[3]

The crucial point to be proved is that, on the one hand, $f(n)$ cannot have an asymptotic distribution function distinct from the infinite convolution which is formally associated with $f(n)$; and that, on the other hand, the mere convergence of this formal convolution implies the existence of an asymptotic distribution function.

In particular, the theory of divergent infinite convolutions[4] will become applicable to every $f(n)$ which does not possess an asymptotic distribution function.

1. For a given function $g(n)$ defined for $n = 1, 2, \cdots$, let $M\{g(n)\}$ denote the limit of the ratio of $g(1) + g(2) + \cdots + g(n)$ and n, as $n \to \infty$, if this limit exists. If S is an increasing sequence of positive integers, let $S(n)$ denote its characteristic function, i. e., the function defined for $n = 1, 2, \cdots$ by placing $S(n) = 1$ or $S(n) = 0$ according as n is or is not

* Received March 23, 1939.

[1] Cf. P. Erdös, I, *Journal of the London Mathematical Society*, vol. 10 (1935), pp. 120-125; II, *ibid.*, vol. 12 (1937), pp. 7-11; III, *ibid.*, vol. 13 (1938), pp. 119-127. In III, further references are given. The earliest instance of the treatment of a non-trivial case, that of (the logarithm of) the ratio $\phi(n)/n$ belonging to Euler's ϕ-function, is due to I. Schur (cf. I. Schoenberg, *Mathematische Zeitschrift*, vol. 28 (1928), p. 194).

[2] Cf. B. Jessen and A. Wintner, *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88.

[3] Cf. P. Hartman, E. R. van Kampen and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 476-486.

[4] Cf. E. R. van Kampen and A. Wintner, *American Journal of Mathematics*, vol. 59 (1937), pp. 635-654.

contained in the set S. If the average $M\{S(n)\}$ exists, it is called the density of S. Generally, let $\bar{M}\{S(n)\}$ denote the upper limit of the ratio of $S(1) + S(2) + \cdots + S(n)$ and n, as $n \to \infty$.

For a given real-valued function $f = f(n)$ and for any real number x, let S_x denote the set of those n at which $f(n) < x$. If there exists a monotone function $\sigma = \sigma(x)$, $-\infty < x < +\infty$, which satisfies the boundary conditions $\sigma(-\infty) = 0$, $\sigma(+\infty) = 1$ and is such that, at every continuity point x of σ, the limit $M\{S_x(n)\}$ exists and equals $\sigma(x)$, then f is said to have an asymptotic distribution function, σ.

2. For a given function $f(n)$ and for every positive integer k, define a function $f^{(k)}(n)$ by placing

$$f^{(k)}(n) = \begin{cases} 0, & \text{if } n \not\equiv 0 \pmod{p_k}, \\ f(p_k^{\,l}), & \text{if } p_k^{\,l} \mid n \text{ and } p_k^{\,l+1} \nmid n, \end{cases} \tag{1}$$

where p_k denotes the k-th prime number. Put

$$f_k(n) = \sum_{j=1}^{k} f^{(j)}(n). \tag{2}$$

It is clear that a function $f(n)$ is additive if and only if so is each of the functions $f_1(n), f_2(n), \cdots$ and one has

$$f(n) = \sum_{k=1}^{\infty} f^{(k)}(n), \tag{3}$$

(where $f(n) = f_k(n)$ for every $k > h_n$, if h_n is sufficiently large for a fixed n). In fact, the additivity of $f(n)$, i. e., the property that $f(n_1 n_2) = f(n_1) + f(n_2)$ whenever $(n_1, n_2) = 1$, is equivalent to the property that, for arbitrary positive integers $k; j_1, \cdots, j_k$,

$$f\Big(\prod_{i=1}^{k} p_i^{j_i}\Big) = \sum_{i=1}^{k} f(p_i^{j_i}), \text{ while } f(1) = 0. \tag{4}$$

In what follows, $f(n)$ will denote an arbitrarily given real-valued additive function.

3. Consider first the particular case of those functions $f(n)$ which are bounded on the sequence of all prime numbers $p_1, p_2, \cdots$,

$$|f(p)| < \text{const.} \tag{5}$$

Then $f(n)$ cannot have an asymptotic distribution function unless the series

$$\sum \frac{f(p)^2}{p} \text{ is convergent.} \tag{6}$$

In fact, suppose, if possible, that $\sum f(p)^2/p = \infty$. Then, if ω denotes any

real number, a recent theorem [5] states that the density (§ 1) of the sequence of those natural numbers m which satisfy the inequality

$$f(m) < \sum_{p<m} \frac{f(p)}{p} + \omega \cdot \left\{ 2 \sum_{p<m} \frac{f(p)^2}{p} \right\}^{\frac{1}{2}}$$

exists and is represented by the normal distribution function $\pi^{-\frac{1}{2}} \int_{-\infty}^{\omega} \exp -t^2 dt$. This clearly implies that if a denotes any positive number, the density of the sequence of those natural numbers m which satisfy the inequality $|f(m)| < a$ is 0. But then $f(n)$ does not have an asymptotic distribution function, since the boundary conditions of § 1 are violated. This completes the proof.

4. The assumption (5) also implies that $f(n)$ cannot have an asymptotic distribution function unless the partial sums of the series $\sum \frac{f(p)}{p}$ form a bounded sequence of numbers.

In fact, if this sequence were not bounded, it would follow from (6) by a straightforward application of the method of Turán,[6] that there exists for every $\epsilon > 0$ a sufficiently large $C = C_\epsilon > 0$ in such a way that the number of those positive integers m which satisfy the condition

$$-C + \sum_{p<m} \frac{f(p)}{p} < f(m) < C + \sum_{p<m} \frac{f(p)}{p}$$

exceeds $(1-\epsilon)m$ for every sufficiently large m. But then the definitions of § 1 imply that $f(n)$ cannot have an asymptotic distribution function. This contradiction completes the proof.

5. Next, it will be shown that if (5) is satisfied, $f(n)$ cannot have an asymptotic distribution function unless the series

$$\sum \frac{f(p)}{p} \text{ is convergent.} \tag{7}$$

The proof of (7) will be based on the estimate

$$\frac{1}{n} \sum_{m=1}^{n} f(m)^2 < \text{Const.}, \tag{8}$$

which will be proved in § 6. Suppose that (8) has already been established.

It is clear from (8) that for every $\epsilon > 0$ and for every n one has $\sum |f(j)| < \epsilon n$, if the summation index j runs through those positive integers j which satisfy both inequalities $j < n$, $|f(j)| \geqq \epsilon^{-1} \cdot \text{Const.}$ It follows, there-

[5] P. Erdös and M. Kac, *Proceedings of the National Academy*, vol. 25 (1939), pp. 206-207. Better: Am. J. Math 62 (1940), 738-742.

[6] P. Turán, *Journal of the London Mathematical Society*, vol. 11 (1936), pp. 125-133.

fore, from the definitions of § 1 that $f(n)$ cannot have as asymptotic distribution function unless the limit $M\{f(n)\}$ exists. Since $f(n)$ is supposed to have an asymptotic distribution function, and since $M\{f(n)\}$ is defined as the limit of the ratio of $f(1)+\cdots+f(n)$ and n, it follows that if

$$\sum_{m=1}^{n} f(m) - n \sum_{p\leqq n} \frac{f(p)}{p} = o(n)$$

is true, then so is (7). But this $o(n)$-estimate is certainly true in case the additive function $f(n)$ is such that $f(p^k)=f(p)$ holds for every prime p and for every $k \geqq 2$, since in this case

$$\sum_{m=1}^{n} f(m) = \sum_{p\leqq n} \left[\frac{n}{p}\right] f(p) = n \sum_{p\leqq n} \frac{f(p)}{p} + o(n)$$

in virtue of (5). Finally, it is known [7] that the case of any $f(n)$ satisfying (5) may be reduced to the case in which not only (5) holds, but also $f(p^k)=f(p)$.

This completes the proof of (7) on the assumption (8).

6. In order to prove (8), notice first that

$$\sum_{m=1}^{n} f(m)^2 = \sum\sum \left[\frac{n}{pq}\right] f(p)f(q) + \sum_{p\leqq n} \left[\frac{n}{p}\right] f(p)^2,$$

where the summation indices p, q of the double sum run through all pairs of distinct primes. Hence, $\frac{1}{n}\sum_{m=1}^{n} f(m)^2$ cannot exceed

$$\left(\sum_{p<n^{\frac{1}{2}}} \frac{f(p)}{p}\right)^2 + 2 \sum_{n^{\frac{1}{2}}\leqq p\leqq n} \frac{f(p)}{p} \sum_{qp\leqq n} \frac{f(q)}{q} + \frac{1}{n} \sum_{qp\leqq n} |f(p)f(q)| + \sum_{p\leqq n} \frac{f(p)^2}{p}.$$

Since the partial sums of the series $\sum \frac{f(p)}{p}$ are bounded (§ 4), it follows that

$$\frac{1}{n}\sum_{m=1}^{n} f(m)^2 = O(1) + O\left(\sum_{n^{\frac{1}{2}}\leqq p\leqq n} \frac{|f(p)|}{p} O(1)\right) + \frac{1}{n} \sum_{qp\leqq n} |f(p)f(q)| + \sum_{p\leqq n} \frac{f(p)^2}{p}.$$

Consequently, from (5) and (6),

$$\frac{1}{n}\sum_{m=1}^{n} f(m)^2 = O(1) + O\left(\sum_{n^{\frac{1}{2}}\leqq p\leqq n} \frac{1}{p}\right) + \frac{1}{n} O\left(\sum_{qp\leqq n} 1\right) + O(1).$$

This clearly implies (8). Thus, the proof of (7) is now complete.

7. For any $y=f(n)$, define $y^+=f^+(n)$ by placing

(9) $\qquad y^+ = y$ or $y^+ = 1$ according as $|y|<1$ or $|y|\geqq 1$.

---

[7] P. Erdös, *loc. cit.* [1], I, p. 124.

Then an $f(n)$ which satisfies (5) cannot have an asymptotic distribution function unless the series

$$\text{(10)} \qquad \Sigma \frac{f^+(p)}{p} \text{ and } \Sigma \frac{f^+(p)^2}{p} \text{ are convergent.}$$

In fact, the convergence of the second of the series (10) is clear from (6) and (9). On the other hand, (7) and (9) show that the first of the series (10) is certainly convergent if $\Sigma' | f(p) |/p < \infty$, where the summation Σ' ranges over those primes p at which $| f(p) | \geqq 1$. Hence, it is sufficient to assure that $\Sigma' f(p)^2/p < \infty$. But this is clear from (6).

8. It will now be easy to prove the following theorem:

(i). *An additive function $f(n)$ has an asymptotic distribution function if and only if* (10) *is satisfied.*

That (10) is a sufficient condition for the existence of the asymptotic distribution function of $f(n)$, has been proved by Erdös, loc. cit.[1], III. On the other hand the necessity of (10) is established by § 7 for those $f(n)$ which satisfy (5). Thus, (i) will be proved by showing that the result of § 7 holds without the restriction (5) also.

Suppose, therefore, that $f(n)$ has an asymptotic distribution function, and that there does not exist a constant which satisfies (5) for all primes. It may be assumed that there does not exist a number K with the property that the sum of the reciprocal values of those primes p at which $| f(p) | > K$ is a convergent series. In fact, if this series is convergent for some sufficiently large K, then the truth of the statement (10) may be obtained by a repetition of the corresponding considerations, applied loc. cit.[7]

9. Consequently, it may be assumed that there exists an infinite sequence of primes, say $p^{(1)}, p^{(2)}, \cdots$, in such a way that, on the one hand, the sum of the reciprocal values of all these $p^{(j)}$ is a divergent series, while the sum of the squares of these reciprocal values is less than $\frac{1}{4}$, say; and, on the other hand, one has $| f(p^{(j)}) | \to \infty$, as $j \to \infty$. Now, the existence of such a sequence $p^{(1)}, p^{(2)}, \cdots$ is incompatible with the existence of an asymptotic distribution function for $f(n)$.

In fact, if $f(n)$ has an asymptotic distribution function, then there exists, by § 1, a sufficiently large $\lambda > 0$ which has the property that the density of the set of those integers n which satisfy $| f(n) | \leqq \lambda$ exists and does not exceed $\frac{3}{4}$, say. Thus, it is clear from the choice of the sequence $p^{(1)}, p^{(2)}, \cdots$, that the density of the set of those integers n which satisfy the condition $| f(n) | \leqq \lambda$ and are not divisible by the square of any of the primes $p^{(1)}, p^{(2)}, \cdots$, cannot be less than $\frac{1}{2}$. Hence, on choosing the value of λ

fixed, one sees that the sum of the reciprocal values of those integers m which satisfy both inequalities $|f(m)| \leqq \lambda, m \leqq n$ and are not divisible by the square of any of the primes $p^{(1)}, p^{(2)}, \cdots$, must exceed $\frac{1}{2}\log n$ for every sufficiently large n. Now, it is readily seen from the elementary considerations which were used in the proof [8] of the sufficiency of the condition (10), that one can find two integers m, say m_1 and m_2, which satisfy the condition $|f(m)| \leqq \lambda$ and are such that, on the one hand, $m_2 = m_1 d$ for a suitable integer d and, on the other hand, $p^{(l)} \mid d$ for a suitable element $p^{(l)}$ of the sequence $p^{(1)}, p^{(2)}, \cdots$.

Clearly, this implies a contradiction whenever the prime $p^{(l)}$ satisfies the condition $f(p^{(l)}) > 2\lambda$. And this condition may be satisfied by a suitable choice of $p^{(l)}$. In fact, the value of λ has been fixed, while $|f(p^{(j)})| \to \infty$, as $j \to \infty$.

This completes the proof of (i).

10. Let $f(n)$ be an arbitrary additive function, which need not have an asymptotic distribution function. On choosing any fixed $k \geqq 1$, and denoting by $u_1, \cdots, u_k$ real variables each of which varies continuously from $-\infty$ to $+\infty$, one readily verifies from the definitions (1)-(2), that limits

$$M\{\exp i \sum_{j=1}^{k} u_j f^{(j)}(n)\} \quad \text{and} \quad M\{\exp i u_j f^{(j)}(n)\}; \quad j = 1, \cdots, k,$$

where $M\{g(n)\} = \lim (g(1) + \cdots + g(n))/n$, exist uniformly in every fixed $(u_1, \cdots, u_k)$-sphere $u_1^2 + \cdots + u_k^2 < \text{const.}$, and that

$$M\{\exp i \sum_{j=1}^{k} u_j f^{(j)}(n)\} = \prod_{j=1}^{k} M\{\exp i u_j f^{(j)}(n)\}. \tag{11}$$

This means [9] that *the k functions* $f^{(1)}(n), \cdots, f^{(k)}(n)$ *are statistically independent*, and implies,[10] therefore, not only that each of the additive functions (1) and (2) possess an asymptotic distribution function, say $\sigma^{(k)}(x)$ and $\sigma_k(x)$, but also that

$$L(u; \sigma_k) = \prod_{j=1}^{k} L(u; \sigma^{(j)}) \text{ for } -\infty < u < +\infty, \tag{12}$$

$L(u; \rho)$ denoting the Fourier-Stieltjes transform

$$L(u; \rho) = \int_{-\infty}^{+\infty} e^{iux} d\rho(x). \tag{13}$$

In fact, if $g(n)$ has the asymptotic distribution function $\rho(x)$, then

---

[8] P. Erdös, *loc. cit.* [1], III, Proof of Lemma 3.

[9] P. Hartman, E. R. van Kampen and A. Wintner, *loc. cit.* [3], § 8.

[10] *Ibid.*, § 7.

$$M\{\exp iug(n)\} = L(u;\rho)\,; \tag{14}$$

so that (12) follows from (11) by choosing all the u_j equal $(=u)$. And (12) means that the asymptotic distribution function $\sigma_k(x)$ of $f_k(n)$ is the convolution

$$\sigma_k = \sigma^{(1)} * \sigma^{(2)} * \cdots * \sigma^{(k)} \tag{15}$$

of the asymptotic distribution functions $\sigma^{(j)}(x)$ of the terms $f^{(j)}(x)$ of $f_k(x)$, i. e., of the sum (2) which belongs to the given $f(n)$ in virtue of (1). It is understood that the convolution, $\alpha * \beta$, of two distribution functions α, β is defined as the distribution function

$$\alpha(x) * \beta(x) = \int_{-\infty}^{+\infty} \alpha(x-y)\,d\beta(y)\,; \text{ so that } L(u;\alpha*\beta) = L(u;\alpha)L(u;\beta).$$

It is readily verified that the asymptotic distribution function $\sigma^{(k)}(x)$ of the additive function (1) is the step function which has at $x = f(p^m)$ the jump $p^{-m}(1-p^{-1})$, where p denotes the k-th prime number and $m = 0, 1, \cdots$. It is understood that the values $f(p^m)$ belonging to the same $p = p_k$ need not be distinct for distinct m. Thus it is clear from (13) that

$$L(u;\sigma^{(k)}) = (1-p^{-1}) \sum_{m=0}^{\infty} p^{-m} \exp iuf(p^m), \text{ where } p = p_k. \tag{16}$$

10 bis. It should be mentioned that the italicized result of § 10 becomes wrong if statistical independence is meant in the sense of Kac and Steinhaus,[11] instead of the sense defined by Hartman, van Kampen and Wintner.[3] In fact, one can readily choose a sequence of numbers $f^{(1)}(2), f^{(1)}(2^2), f^{(1)}(2^3), \cdots$ in such a way that for the corresponding function $f^{(1)}(n)$, determined by (1), the limit $M\{S_x(n)\}$ mentioned at the end of § 1 does not exist at $x = 0$, say.

11. Let the additive function $f(n)$, the positive integer k and an $\epsilon > 0$ be arbitrarily given. Let S_k^{ϵ} denote the set of those positive integers n at which the function (2) satisfies the inequality $|f(n) - f_k(n)| > \epsilon$. Suppose that, in the notations introduced in § 1, one has $\bar{M}\{S_k^{\epsilon}(n)\} \to 0$, as $k \to \infty$. Suppose finally that this limit relation holds for every fixed $\epsilon > 0$. Then the functions $f_1(n), f_2(n), \cdots$ are said to tend to $f(n)$ in relative measure.

The following theorem will now be proved:

(ii). *An additive function $f(n)$ has an asymptotic distribution if and only if the function (2) tends, as $k \to \infty$, to $f(n)$ in relative measure, in which*

[11] M. Kac and H. Steinhaus, *Studia Mathematica*, vol. 7 (1937), pp. 1-15.

case the asymptotic distribution function of $f_k(n)$ tends, as $k \to \infty$, to the asymptotic distribution function of $f(n)$.

In fact, if $f_k(n)$ tends to $f(n)$ in relative measure, then $f(n)$ has an asymptotic distribution function which is the limit of the asymptotic distribution function of $f_k(n)$. This follows from a general theorem,[12] since every $f_k(n)$ has, by § 10, an asymptotic distribution function. Conversely, if $f(n)$ has an asymptotic distribution function, then (i), § 8, shows that (10) is satisfied. Since (10) is known [13] to be sufficient for the convergence of $f_k(n)$ to $f(n)$ in relative measure, the proof of (ii) is complete.

12. On comparing the definition (16) of $\sigma^{(k)}(x)$ with a general criterion [14] for the convergence of an arbitrary infinite convolution, one readily sees that (10) is precisely the condition for the convergence of the infinite convolution

$$\sigma = \sigma^{(1)} * \sigma^{(2)} * \cdots . \tag{17}$$

But the k-th approximating convolution of (17) is the asymptotic distribution function (15) of $f_k(n)$. Hence, (i) and (ii) imply the following theorem:

(iii). *An additive function $f(n)$ has an asymptotic distribution if and only if the infinite convolution* (17) *is convergent. Furthermore, if $f(n)$ has at all an asymptotic distribution function, the latter is necessarily* (17).

13. As pointed out at the end of § 10, the distribution function $\sigma^{(k)}(x)$ is a step function. Hence, on combining (iii) with a general result [15] on infinite convolutions of step functions, one obtains the following theorem:

(iv). *A distribution function which is the asymptotic distribution function of an additive function $f(n)$ either is a step function or it is everywhere continuous, and in the latter case it is either purely singular or absolutely continuous.*

It is a rather intricate problem, at least in general, to distinguish between the cases of absolutely continuous and purely singular behavior.[16] On the other hand, the case of mere continuity is completely characterized by the following theorem: [17]

---

[12] B. Jessen and A. Wintner, *loc. cit.* [2], Theorem 24.

[13] P. Erdös, *loc. cit.* [1], III, Lemma 1.

[14] B. Jessen and A. Wintner, *loc. cit.* [2], Theorem 34.

[15] B. Jessen and A. Wintner, *loc. cit.* [2], Theorem 35.

[16] There are two extreme cases in which such a distinction is not too difficult; cf. P. Erdös, *American Journal of Mathematics*, vol. 61 (1939), pp. 722-725.

[17] If use is made of (i), the criterion (18) also follows as a consequence of the results obtained by Erdös, *loc. cit.* [1], III, by direct considerations.

(v). *A distribution function which is the asymptotic distribution function of an additive function $f(n)$ has no discontinuities if and only if*

$$\sum_{f(p)\neq 0} \frac{1}{p} \text{ is divergent.} \tag{18}$$

In fact, a general theorem of P. Lévy [18] states that a convergent infinite convolution $\rho^{(1)} * \rho^{(2)} * \cdots$ of distribution functions $\rho^{(k)}$ is continuous at every x if and only if the infinite product $\Pi d_k = 0$, where d_k denotes the largest jump $(0 \leqq d_k \leqq 1)$ of $\rho^{(k)}(x)$. On the other hand, (16) shows that the largest jump of $\sigma^{(k)}(x)$ is at most $1 - p^{-1} + p^{-2}$ if $f(p) \neq 0$, and at least $1 - p^{-2}$ if $f(p) = 0$, where $p = p_k$. Hence, (v) is implied by (iii).

14. By the spectrum of a distribution function $\rho = \rho(x)$, $-\infty < x < +\infty$, is meant the set of those x in the vicinity of which ρ is not constant, i. e., of those x for which one has $\rho(x-\epsilon) < \rho(x+\epsilon)$ whenever $\epsilon > 0$.

(vi). *If the additive function $f(n)$ has an asymptotic distribution function, the spectrum of the latter is identical with the closure set of the values attained by the function $f(n)$ for $n = 1, 2, \cdots$.*

In fact, (16) shows that (vi) is certainly true in the particular case $f(n) = f^{(k)}(n)$. Hence, on combining (iii) with a general theorem,[19] one sees from (3)-(4), where $f^{(k)}(1) = 0$, that (vi) is true for an arbitrary $f(n)$.

15. Further results may be obtained by combining (iii) with the theory of divergent infinite convolutions.[20]

THE INSTITUTE OF ADVANCED STUDY,
THE JOHNS HOPKINS UNIVERSITY.

---

[18] P. Lévy, *Studia Mathematica*, vol. 3 (1931), pp. 119-155, Théorème XIII. Another proof of Lévy's theorem has recently been communicated to one of us by Professor B. Jessen.

For a unified treatment of Lévy's theorem and of all of the convolution theorems used above, cf. a paper by E. R. van Kampen, which will appear in the *American Journal of Mathematics*.

[19] B. Jessen and A. Wintner, *loc. cit.* [2], Theorem 3.

[20] E. R. van Kampen and A. Wintner, *loc. cit.* [4], Theorems 4-9, and the investigations of P. Lévy, referred to there.

ON THE SMOOTHNESS OF THE ASYMPTOTIC DISTRIBUTION OF ADDITIVE ARITHMETICAL FUNCTIONS.*

By PAUL ERDÖS.

Introduction. Starting with any given sequence $a_2, a_3, \cdots, a_p, \cdots$ of real numbers, define a sequence $f_1, f_2, f_3, \cdots$ by placing $f_n = \Sigma a_p$, where the summation runs through all prime divisors p of n (in particular, $f_1 = 0$). Clearly, $f_{n+m} = f_n + f_m$ whenever $(n, m) = 1$.

Put $a^+{}_p = a_p$ or $a^+{}_p = 1$ according as $-1 < a_p < 1$ does or does not hold. It is known [1] that the additive function f_n of n possesses an asymptotic distribution function $\sigma(x), -\infty < x < +\infty$, if and only if the series

$$\Sigma \frac{a^+{}_p}{p} \text{ and } \Sigma \frac{(a^+{}_p)^2}{p} \text{ are convergent,} \tag{1}$$

in which case the Fourier-Stieltjes transform

$$L(u) = \int_{-\infty}^{+\infty} e^{iux} d\sigma(x), \qquad -\infty < u < +\infty \tag{2}$$

is represented for every real u by the convergent product

$$L(u) = \Pi \left(1 - \frac{1 - \exp(ia_p u)}{p}\right). \tag{3}$$

The relation (3) and a general theorem of P. Lévy imply [2] that the distribution function $\sigma(x)$ is continuous for $-\infty < x < +\infty$ if and only if

$$\sum_{a_p \neq 0} \frac{1}{p} = \infty. \tag{4}$$

It will always be assumed that (1) and (4) are satisfied.

It follows from a general theorem of Jessen and Wintner [3] that the monotone continuous function $\sigma(x)$ either is absolutely continuous or purely singular for $-\infty < x < +\infty$. The object of the present note is to show that either of these cases can actually occur for additive arithmetical functions f_n of simple type.

* Received March 23, 1939.

[1] P. Erdös and A. Wintner, *American Journal of Mathematics*, vol. 61 (1939), pp. 713-721.

[2] P. Lévy, *Studia Mathematica*, vol. 3 (1931), p. 150; cf. *loc. cit.* [1]

[3] B. Jessen and A. Wintner, *Transactions of the American Mathematical Society*, vol. 39 (1935), p. 86; cf. *loc. cit.* [1].

In particular, the result of § 1 will imply that if

$$a_p = \frac{(-1)^{\frac{1}{2}(p-1)}}{(\log\log p)^{3/4}}, \qquad (p > e^e), \tag{5}$$

then there exists a transcendental entire function $\sigma(z) = \sigma(x+iy)$ which reduces for $y=0$ to the distribution function of f_n. On the other hand, the result of § 3 will show that if

$$f_n = \log \frac{n}{\phi(n)}, \tag{6}$$

where $\phi(n)$ is Euler's function, then the distribution function of f_n is not absolutely continuous.

1. The method of this § 1 is, in contrast to that of § 3, not of an elementary nature, and consists of an adaptation of a method applied by Wintner to Bernoulli convolutions and the corresponding distribution functions occurring in the theory of the Riemann zeta function.[4] This method consists in estimating the product (3) for large $|u|$ by the following approach: Since each of the factors of (3) has, for every u, an absolute value not exceeding 1, it is clear that

$$L(u) \leqq \prod_{A(u)<p<B(u)} \left| 1 - \frac{1-\exp(ia_p u)}{p} \right| \tag{7}$$

holds for arbitrary positive functions $A(u)$, $B(u)$ of u. And the method consists in choosing $A(u)$, $B(u)$, if possible, in such a way as to assure that

$$L(u) = O(\exp - C|u|^c), \qquad u \to \pm\infty, \tag{8}$$

holds for some pair of positive constants c, C. If (2) satisfies (8), then $\sigma(x)$ has for every x derivatives of arbitrary high order; while if (8) holds for $c=1$ and some $C>0$, then $\sigma(x)$ is regular analytic and bounded in every strip $|\Im x| < \text{const.} < C$ about the real axis of the complex x-plane.[5] In particular, $\sigma(x)$ is an entire function if (8) holds for some $c>1$ and for some $C>0$.

Suppose that $|a_p|$ is monotone in p, and define the range of p-values over which the product (7) is extended by

$$A(u) < p < B(u): \quad 3\pi < 4|a_p u| < 5\pi. \tag{9}$$

Then each factor of the product on the right of (7) is less than $1-1/p$; so that $L(u) < \Pi'(1-1/p)$, where p runs, for every fixed u, over the range (9).

---

[4] A. Wintner, *Bulletin of the American Mathematical Society*, vol. 41 (1935), pp. 137-138; *American Journal of Mathematics*, vol. 61 (1939), pp. 231-236.

[5] A. Wintner, *American Journal of Mathematics*, vol. 56 (1934), p. 659.

Hence, by Mertens' elementary result $\prod_{p<t}(1-1/p) \sim e^{-\gamma}/\log t$,

$$L(u) = O\left(\frac{\log A(u)}{\log B(u)}\right), \qquad u \to \pm\infty, \tag{10}$$

where $A(u)$, $B(u)$ are defined by (9).

For instance, if a_p is given by (5), then (8) is satisfied by a $c > 1$ and a $\delta' > 0$. In fact, if the exponent $\frac{3}{4}$ of the denominator of (5) is replaced by an arbitrary $\alpha > \frac{1}{2}$, then (10) and (9) clearly imply that (5) is satisfied by $c = 1/\alpha$ and $C > 0$. Hence, $\sigma(x)$ is an entire function if $\frac{1}{2} < \alpha < 1$; it is regular analytic at least in a strip $|\Im x| <$ const. if $\alpha = 1$; and it has, at least, derivatives of arbitrarily high order for every u if $\alpha > 1$. It may be mentioned that if $\alpha > 1$, the distribution function $\sigma(x)$ has derivatives of arbitrarily high order also when one omits in (5) the factor $(-1)^{\frac{1}{2}(p-1)}$; in which case $\sigma(x)$ cannot be regular analytic along the real axis, since $\sigma(x) = 0$ for every $x < 0$.

2. Let $a_p = 2^{-p}$. Then it is readily verified from (3) that $L(2^m\pi)$ tends, as $m \to \infty$, to a positive limit; so that

$$L(u) \to 0, \qquad u \to \pm\infty, \tag{11}$$

does not hold. It follows, therefore, from the extension of the Riemann-Lebesgue lemma to (2), that the distribution function $\sigma(x)$ is singular.

Of course, (11) only is a necessary condition for the absolute continuity of $\sigma(x)$; in fact,[6] not even $L(u) = O(u^{-\frac{1}{2}+\epsilon})$, where $\epsilon > 0$ is arbitrarily small, is capable of assuring the absolute continuity of $\sigma(x)$.

On the other hand, it is clear from Plancherel's theorem that if $L(u) = O(u^{-\frac{1}{2}-\epsilon})$ holds for some $\epsilon > 0$, then $\sigma(x)$ must be absolutely continuous. This estimate of $L(u)$ is satisfied in case $a_p = 1/\log p$. In order to see this, one merely needs a slight improvement on the crude step (7) and repeated application of Mertens' asymptotic formula, used in § 1.

3. In contrast to the result of § 1, it will now be shown that if

$$a_p = O(p^{-c}), \qquad p \to \infty, \tag{12}$$

holds for some $c > 0$, then $\sigma(x)$ is singular.

In the proof two elementary facts will be needed:

(I) Choosing a fixed large N, write every positive integer m in the form $m = m'm''$, where m' is composed of primes $\leqq N$, and m'' of primes $> N$. Then the density of those m which satisfy the inequality $m' < N^{c/4}$ exceeds a positive lower bound a which depends on $c > 0$ but not on N.

In fact, the density of the positive integers which are not divisible by

[6] Cf. N. Wiener, and A. Wintner, *American Journal of Mathematics*, vol. 60 (1938), pp. 513-522.

any prime $\leqq N$ is $\Pi(1-1/p)$, where $p \leqq N$ (sieve of Eratosthenes). Thus, it is readily seen that the density of the integers $m = m'm''$ for which $m' < N^{c/4}$ is

$$\sum_{m' < N^{c/4}} \frac{1}{m'} \prod_{p \leqq N} \left(1 - \frac{1}{p}\right).$$

Hence, (I) follows from $\prod_{p \leqq N} (1 - 1/p) \sim e^{-\gamma}/\log N$.

(II) For a fixed large N and for $k = 1, 2, \cdots$, put $g_k = \Sigma a_p$, where the summation runs through those prime divisors p of k which do not exceed N, and the a_p satisfy (12). Let f_k be defined, as in the Introduction, by the sum Σa_q, where q runs through the prime divisors q of k. Then there exists a $b > 0$ which is independent of N and has the property that the density of those positive integers k which satisfy the inequality $|f_k - g_k| > N^{-c/2}$ cannot exceed $bN^{-c/2}$.

In fact, it is clear that, for an arbitrary n,

$$\sum_{k=1}^{n} |f_k - g_k| \leqq \sum_{p > N} |a_p| = O\left(n \sum_{p > N} \frac{1}{p^{1+c}}\right) < \frac{bn}{N^c},$$

where b is a constant. Thus, the density of those positive integers k which satisfy both inequalities $k \leqq n$, $|f_k - g_k| > N^{c/2}$ cannot be greater than $bnN^{-c/2}$. This clearly implies (II).

It is now easy to show [7] that $\sigma(x)$ is singular on the assumption (12). In fact, let N be large. Consider the x-intervals

$$f_m - N^{-c/2} < x < f_m + N^{c/2}, \text{ where } m = 1, 2, \cdots, [N^{c/4}].$$

It follows from (I) and (II) that the density of those positive integers k for which $x = f_k$ lies in one of these $[N^{c/4}]$ intervals cannot be less than $a - bN^{-c/2}$ and exceeds, therefore, a fixed lower bound $C > 0$ for all sufficiently large N. And the sum of the lengths of these $[N^{c/4}]$ intervals is majorized by $N^{-c/4}$. Hence, on letting $N \to \infty$, one sees from the definition of $\sigma(x)$ as the asymptotic distribution function of f_n, that $\sigma(x)$ cannot be absolutely continuous.

Since (12) implies (4), it follows from the general theorem of Jessen and Wintner, referred to in the Introduction, that $\sigma(x)$ is purely singular. It may be mentioned that, in the particular case (12), a direct and elementary discussion could also assure that $\sigma(x)$ does not possess an absolutely continuous component.

THE INSTITUTE FOR ADVANCED STUDY.

---

[7] For a similar argument, cf. E. R. van Kampen and A. Wintner, *Journal of the London Mathematical Society*, vol. 12 (1937), pp. 243-244; also P. Hartman and R. Kershner, *American Journal of Mathematics*, vol. 59 (1937), pp. 809-822.

ON A CHARACTERIZATION OF THE NORMAL DISTRIBUTION.*

By M. KAC.**

The object of the present note is to give a simple characterization of the normal distribution. This characterization is based on an invariance principle and admits a physical interpretation.

1. Let $f(\alpha)$ be a measurable function defined in $(0, 1)$. By the distribution function $\sigma_f(\omega)$ of the function $f(\alpha)$ one understands the measure of the set of those α's for which $f(\alpha) < \omega$.

THEOREM. *If for every γ the functions*

$$f_\gamma(\alpha) = \cos\gamma f(\alpha) + \sin\gamma g(\alpha), \quad g_\gamma(\alpha) = \sin\gamma f(\alpha) - \cos\gamma g(\alpha)$$

are statistically independent,[1] then the distribution functions of $f(\alpha)$ and $g(\alpha)$ are normal and symmetric with the same precision h,

$$\sigma_f(\omega) = \sigma_g(\omega) = h\pi^{-\frac{1}{2}} \int_{-\infty}^{\omega} \exp(-h^2u^2)\,du.$$

(*The case $h = \infty$ of the distribution function $\frac{1}{2}(1 + \operatorname{sign}\omega)$ is included.*)

Let

$$\int_0^1 \exp(i\xi f(\alpha))\,d\alpha = A(\xi), \qquad \int_0^1 \exp(i\xi g(\alpha))\,d\alpha = B(\xi)$$

and let us suppose that $A(\xi) = A(-\xi)$ and $B(\xi) = B(-\xi)$. This preliminary assumption is to the effect that the distributions are symmetric, and have, therefore real Fourier transforms $A(\xi)$, $B(\xi)$. The independence of $f_\gamma(\alpha)$ and $g_\gamma(\alpha)$ implies, for arbitrary real ξ and η, that

$$\text{(i)} \quad \int_0^1 \exp i[\xi f_\gamma(\alpha) + \eta g_\gamma(\alpha)]\,d\alpha = \int_0^1 \exp(i\xi f_\gamma(\alpha))\,d\alpha \cdot \int_0^1 \exp(i\eta g_\gamma(\alpha))\,d\alpha$$
$$= A(\xi\cos\gamma)\,B(\xi\sin\gamma)\,A(\eta\sin\gamma)\,B(-\eta\cos\gamma).$$

* Received December 23, 1938; Revised April 21, 1939.

** Fellow of the Parnas Foundation, Lwów, Poland.

[1] Cf. for instance, M. Kac, "Sur les fonctions indépendantes," I, *Studia Mathematica*, vol. 6 (1936), pp. 46-58.

On the other hand,

$$\xi f_\gamma(\alpha) + \eta g_\gamma(\alpha) = (\xi \cos\gamma + \eta \sin\gamma) f(\alpha) + (\xi \sin\gamma - \eta \cos\gamma) g(\alpha),$$

and therefore

$$\text{(ii)} \quad \int_0^1 \exp i[\xi f_\gamma(\alpha) + \eta g_\gamma(\alpha)]\, d\alpha = A(\xi \cos\gamma + \eta \sin\gamma) B(\xi \sin\gamma - \eta \cos\gamma).$$

Comparison of (i) with (ii) gives

$$\text{(iii)} \quad A(\xi \cos\gamma + \eta \sin\gamma) B(\xi \sin\gamma - \eta \cos\gamma) = A(\xi \cos\gamma) A(\eta \sin\gamma) B(\xi \sin\gamma) B(-\eta \cos\gamma).$$

Putting successively $\gamma = \pi/4$ and $\gamma = 3\pi/4$ and writing ξ and η instead of $\xi/\sqrt{2}$ and $\eta/\sqrt{2}$, one obtains

$$A(\xi + \eta) B(\xi - \eta) = A(\xi) A(\eta) B(\xi) B(\eta),$$
$$A(\eta - \xi) B(\xi + \eta) = A(\xi) A(\eta) B(\xi) B(\eta),$$

since $A(\xi) = A(-\xi)$, $B(\xi) = B(-\xi)$ by assumption. Consequently,

$$A(\xi + \eta) B(\xi - \eta) = A(\eta - \xi) B(\xi + \eta).$$

Placing $\eta = \xi$, it follows that

$$A(2\xi) = B(2\xi).$$

Thus, the problem is reduced to the functional equation

$$A(\xi + \eta) A(\xi - \eta) = A^2(\xi) A^2(\eta).$$

In particular, $A(2\xi) = A^4(\xi)$; so that, since $A(\xi)$ is real, $A(\xi) \geqq 0$. Hence, repeated application of $A(2\xi) = A^4(\xi)$ gives

$$A(\xi/2^k) = (A(\xi))^{1/4^k}$$

As $A(\xi/2^k) \to 1$ if $k \to \infty$, it follows from the continuity of $A(\xi)$ that $A(\xi) > 0$ for every ξ. Putting successively $\eta = \xi, 2\xi, 3\xi, \cdots$ and applying a well known method of Cauchy, it follows from $A(\xi) > 0$ in a few steps that

$$A(p\xi/q) = A^{p^2/q^2}(\xi)$$

for arbitrary integers p, q; so that, since $A(\xi)$ is continuous,

$$A(\xi) = \exp(k\xi^2) \qquad (\exp k = A(1)).$$

As $0 < A(\xi) \leq 1$, one has $k \leq 0$, and so, if $k = -h^2/4$, the proof is complete in the symmetric case under consideration.

2. According to a remark of Professor Wintner, the assumptions $A(\xi) = A(-\xi)$ and $B(\xi) = B(-\xi)$ of the symmetric case are superfluous. In fact, let

$$A^*(\xi) = A(\xi)A(-\xi) \quad \text{and} \quad B^*(\xi) = B(\xi)B(-\xi).$$

Then the equation (iii) holds for A^* and B^*, instead of for A and B. Hence,

$$A^*(\xi) = \exp(k\xi^2).$$

Applying a well-known theorem of Cramér,[2] it follows that

$$A(\xi) = e^{a\xi^2+b_1\xi}, \quad B(\xi) = e^{a\xi^2+b_2\xi}.$$

Finally, an easy calculation shows that $b_1 = b_2 = 0$. This completes the proof of the theorem in the general case.

3. If instead of functions defined in $(0, 1)$ one considers independent functions defined in $(-\infty, +\infty)$ and the independence is meant in the sense of Hartman, van Kampen and Wintner,[3] one sees that the theorem remains unchanged (it is of course understood that in this case one deals with asymptotic distribution functions).

It is also clear that the theorem of § 1 holds in more than two dimensions, in the sense that instead of the two-dimensional orthogonal matrix one has to consider a more-dimensional one. The theorem in its most general form may then be stated as follows:

If a n-dimensional vector function $V(t)$ is such that its components are statistically independent in every system of Cartesian coördinates, then the density of the asymptotic distribution function of $V(t)$ is Gaussian and of radial symmetry.

If $V(t)$ is the velocity of a gas molecule $(n = 3)$ then the theorem gives the well known result of Maxwell in its "time average" formulation. It may be mentioned that Maxwell assumed not only the independence but also the radial symmetry of the distribution function. In our case the theorem follows only from the invariant character of the independence.

THE JOHNS HOPKINS UNIVERSITY.

---

[2] H. Cramér, "Random variables," *Cambridge Tracts*, 1937, p. 52.

[3] P. Hartman, E. R. van Kampen and A. Wintner, "Asymptotic distributions and statistical independence," *American Journal of Mathematics*, vol. 61 (1939), pp. 477-486.

ON THE ASYMPTOTIC DISTRIBUTION OF A UNIFORMLY ALMOST PERIODIC FUNCTION.*

By E. R. VAN KAMPEN.

For a given real-valued, uniformly almost-periodic function $f(t)$, $-\infty < t < +\infty$, and for any real number x, let A_x; B_x; C_x denote the t-sets on which $f(t) \leqq x$; $f(t) < x$; $f(t) = x$, respectively. According to Wintner,[1] every $f(t)$ has an asymptotic distribution function $\sigma = \sigma(x)$, $-\infty < x < +\infty$, in the following sense: If x is a continuity point of the monotone function σ, then any of the three sets A_x; B_x; C_x has a relative measure, represented by $\sigma(x)$; $\sigma(x)$; 0, respectively. The same does not need to hold if x is a discontinuity point of σ. In fact, Bohr [2] has constructed a real, non-negative, uniformly almost periodic function $f(t)$ for which the set $A_0 = C_0$ does not possess a relative measure.

The question was raised by Wintner whether Bohr's example, or a suitable modification, is susceptible to the method of condensation of singularities. The object of this note is to prove that if $x_1, x_2, \cdots$ is a given bounded sequence of distinct x-values, then an almost periodic function may be constructed with the property that C_x does not have a relative measure for any $x = x_k$ in the given sequence, and does have a relative measure for any x not in this sequence. For the $x = x_k$, it is, of course, impossible that both sets A_x, B_x have a relative measure.

The following application of Bohr's construction leads to a class of functions $h_a(x)$ for which C_0 does not have a relative density and which is sufficiently flexible as to allow a process of condensation.

Let a be a number between 0 and 1 and let $\phi(t) = \phi(t+1)$ be a continuous periodic function defined in such a way that $\phi(t) = 0$ for $0 \leqq t \leqq a$, that $0 < \phi(t) \leqq 1$ for $a < t < 1$, and that, for every x, the equation $\phi(t) = x$ has at most a finite number of solutions in the interval $a < t < 1$. Let the positive integers q_n, p_n be such that $\Pi(1 - 1/q_n) > \frac{1}{2}(a+1) > a$, $p_n = p_{n-1}q_n$, $p_0 = 1$ and let the sequence of positive numbers c_n be such that $\Sigma c_n \leqq 1$.

* Received April 20, 1939.

[1] A. Wintner, *Zeitschrift für Physik*, vol. 48 (1928), pp. 148-161; *Mathematische Zeitschrift*, vol. 30 (1929), pp. 290-319.

[2] H. Bohr, "Danske Videnskabernes Selskab," *Mathematiske-Fysiske Meddelelser*, vol. 10, Nr. 6 (1930), pp. 12-17.

The function $h_a(x)$ will be defined as a series $\Sigma c_n \psi_n(t)$, where the $\psi_n(t)$ are continuous periodic functions and $0 \leqq \psi_n(t) \leqq 1$. Thus $h_a(t)$ will be a uniformly almost periodic function, and $0 \leqq h_a(t) \leqq 1$. The functions $\psi_n(t)$ are defined as follows:

$$\psi_n(t) = 0 \text{ for } 0 \leqq t \leqq p_{n-1} \text{ and } 2p_{n-1} \leqq t \leqq p_n;$$
$$\psi_n(t) = \phi(t) \text{ for } p_{n-1} \leqq t \leqq 2p_{n-1}; \qquad \psi_n(t + p_n) = \psi_n(t).$$

If C_0 is the set of t-values for which $h_a(t) = 0$, it will be necessary to determine the ratio of the measures of the common part of C_0 with an interval and of the whole interval. Let α_n and β_n denote these ratios for the intervals $0 \leqq t \leqq p_n$ and $0 \leqq t \leqq 2p_{n-1}$, respectively. Since $\psi_{n+k}(t) = 0$ for $0 \leqq t \leqq p_n$, $k > 0$, these ratios α_n and β_n only depend on the first n terms of the series defining $h_a(t)$. The above-mentioned ratio for the interval $p_{n-1} \leqq t \leqq 2p_{n-1}$ is the same as the corresponding ratio for the function $\phi(t)$, and is therefore a. Since $\psi_k(t)$, $k < n$, has the period p_{n-1}, and $p_n = p_{n-1}q_n$, it follows that

$$\beta_n = \tfrac{1}{2}(\alpha_{n-1} + a) \quad \text{and} \quad \alpha_n = (1 - 1/q_n)\alpha_{n-1} + a/q_n.$$

Clearly, the relative measure of C_0 does not exist if the α_n and β_n do not have equal limits, as $n \to \infty$. By the first equality, this is the case if α_n does not tend to the limit a. But the second equality obviously implies $\alpha_n > \prod_{k=1}^{n} (1 - 1/q_k)$; so that, from the assumption concerning q_n, one has $\alpha_n > \frac{1}{2}(a + 1) > a$. For any $x > 0$, the sets C_x corresponding to the function $h_a(t)$ are at most enumerable sets, since $h_a(t)$ is equal to a constant multiple of $\phi(t)$ in every interval $m \leqq t \leqq m + 1$, $m = 0, \pm 1, \pm 2, \cdots$.

Thus, the function $h_a(t)$, $0 < a < 1$, has the following properties:

(i) $h_a(t)$ is uniformly almost periodic and $0 \leqq h_a(t) \leqq 1$;

(ii) $h_a(t) = 0$ if the fractional part of t is between 0 and a;

(iii) the t-set C_0 defined by $h_a(t) = 0$ does not have a relative measure on $-\infty < t < +\infty$;

(iv) the sets C_x belonging to $h_a(t)$ are at most enumerable for every $x \neq 0$.

Besides $h_a(t)$, use will be made of another auxiliary function, $\chi(t)$, with the properties:

(v) $\chi(t)$ is continuous, $\chi(t + 1) = \chi(t)$ and $\chi(0) = 0$;

(vi) the equation $\chi(t) = x_k$ is satisfied on an interval $\gamma_k \leqq t \leqq \delta_k$, where $0 < \gamma_k < \delta_k < 1$;

(vii) if the solutions in (vi) are excluded, any equation $\chi(t) = x$ has at most an enumerable set of solutions.

The almost periodic function, $f(t)$, with the property that the corresponding sets C_x do not have a relative measure if $x = x_k$ belongs to the given sequence $x_1, x_2, \cdots$, will now be constructed as follows:

The functions $\xi_k(t) = h_a(t - \delta_k)$, where $a = 1 + \gamma_k - \delta_k$, are almost periodic. Furthermore (ii) implies that $\xi_k(t) = 0$ if t is not congruent (mod 1) to any number contained in the interval $\gamma_k \leqq t \leqq \delta_k$. Now

$$f(t) = \chi(t) + \sum_k d_k \xi_k(t)$$

has the desired property.

In fact, that $f(t)$ is uniformly almost periodic, is clear from (i), (v) and the uniform convergence of the series. If k is fixed, let C^k be the t-set defined by $f(t) = x_k$. By (iv), (vi) and (vii), the set C^k consists of an at most enumerable set and of those intervals $\gamma_k \leqq t - n \leqq \delta_k$ for which $\xi_k(t) = 0$. But the set formed by these intervals does not have a relative measure, by (ii) and (iii); so that C^k does not have a relative measure for any k. On the other hand, if x is not in the given sequence $x_1, x_2, \cdots$, then the set C_x defined by $f(t) = x$ is at most enumerable, by (iv), (vi) and (vii). Thus these C_x have the relative measure 0 and $f(t)$ has all properties announced in the introduction. It can also be proved, that the sets A_x, B_x corresponding to this $f(t)$ have a relative measure for any x not in the sequence $x_1, x_2, \cdots$.

A modification, $g(t)$, of the above example, $f(t)$, was communicated to me by Dr. M. Kac. This $g(t)$ is constructed as follows:

Let $h_a(t)$ be defined as above and put $H(m) = h_a(m + \frac{1}{2}(a + 1))$, for $m = 0, \pm 1, \pm 2, \cdots$. Then $H(m)$ is a uniformly almost periodic sequence of non-negative numbers with the property that the set of integers m for which $H(m) = 0$ does not have a density.

Let $x_1, x_2, \cdots$ be the given bounded sequence of distinct positive numbers and let $c > 0$ be such that the equality $cx_l H(m) + x_l = x_k$ is not satisfied for any l, k, m. Now, define $g(t)$ by placing $g(t) = (cH(m) + 1)\chi(t)$, if $m \leqq t < m + 1$.

This $g(t)$ is continuous, by (i). It follows that $g(t)$ is uniformly almost periodic, since so is the sequence $H(m)$. As a consequence of the inequalities

$cx_l H(m) + x_l \neq x_k$ and the character of $\chi(t)$, the t-set C^k defined by $g(t) = x_k$ consists of those intervals $\gamma_k \leqq t - m \leqq \delta_k$ for which $H(m) = 0$ and of at most a sequence of points. Since the set of integers m for which $H(m) = 0$ does not have a density, it follows that the t-set C^k does not have a relative measure.

Notice that the above examples are limit periodic (grenzperiodisch) functions. Apparently it is considerably more difficult to find examples of functions with the same type of singularities but which are not limit periodic.

THE JOHNS HOPKINS UNIVERSITY.

SEMI-LOCALLY CONNECTED SETS.*

By G. T. WHYBURN.

Since so many fundamental propositions in topology have been shown to hold only for locally connected continua whereas their validity for more general classes of sets is at least possible and in many cases probable, the desirability of identifying and studying classes of continua which include the locally connected ones but do not include all continua has long been recognized. It is toward this end that the efforts made in the present paper are directed. The sets studied are defined in such a way that certain types of "oscillation" are eliminated, whereas other, possibly simpler, types are permitted. Since the sets obtained include all compact locally connected continua and also other continua not locally connected, it seems appropriate to call them *semi-locally connected*.[1]

The paper is divided into three parts. The first of these deals with certain fundamental considerations closely related to the definition and contains, among other results, the rather unusual product theorem to the effect that if either of two connected metric sets is semi-locally-connected, so also is their topological product. In the second, the structure of a semi-locally-connected compact continuum is developed from the standpoint of its cyclic element decomposition. It is found that the structure from this viewpoint is so similar to the locally connected case that it may be said that any differences which occur must take place within the individual cyclic elements. The results in this direction culminate in the theorem that any compact semi-locally-connected continuum M can be topologically imbedded in a compact locally connected continuum L so that the intersection of M by any cyclic element of L is exactly a cyclic element of M. It also follows from results in this part that the classic theorem asserting that any locally connected continuum irreducible between two points is a simple arc joining these two points may be extended to hold for semi-locally-connected continua. In the third part, plane semi-locally-connected continua are considered. It is shown that

* Received February 14, 1939.

[1] This term should not be confused with the term *semi-connected* as used by R. L. Wilder in the *Duke Mathematical Journal*, vol. 1 (1935), p. 549. Compare, however, with the notions introduced by Wilder in his paper "Sets which satisfy certain avoidability conditions," *Časopis pro Pěstování Matematiky a Fysiky*, vol. 67 (1938), pp. 185-198.

the boundary of each complementary domain of such a continuum is itself a locally connected continuum. Thus the classic theorem on plane locally connected continua proven in 1921 by Marie Torhorst [2] is substantially extended. It follows, incidentally, that although a semi-locally-connected continuum is not necessarily arcwise connected (see § 2), any such continuum in a plane must contain simple arcs, since the boundary of each of its complementary domains is locally connected.

I. Definitions, Examples, Preliminary Results.

1. Definitions. A connected metric space M will be said to be *semi-locally-connected* at a point p provided that for each $\epsilon > 0$ a $\delta > 0$ exists such that $M - V_\epsilon(p)$ is contained in a finite number of components of $M - V_\delta(p)$. (Note: In general, $V_r(X)$ will denote the r-neighborhood of the set X, i. e., the set of all points x such that the distance $\rho(x, X)$ from x to X is less than r.) If M is semi-locally-connected at each of its points, M is said to be semi-locally-connected.

For the case where M is cyclic, (i. e., without cut points) it follows from (6. 22) below that the property of semi-local-connectivity stands in a sort of dual relationship to local connectivity which may be illustrated as follows:

A cyclic connected metric space M is $\left\{\begin{matrix} \textit{locally connected} \\ \textit{semi-locally-connected} \end{matrix}\right\}$ *at a point $p \in M$ if and only if for each $\epsilon > 0$ a $\delta > 0$ exists such that any two points* $\left\{\begin{matrix} \textit{inside } V_\delta(p) \\ \textit{outside } V_\epsilon(p) \end{matrix}\right\}$ *can be joined by a connected set lying* $\left\{\begin{matrix} \textit{inside } V_\epsilon(p) \\ \textit{outside } V_\delta(p) \end{matrix}\right\}$.

2. Examples. (i) Any compact locally connected continuum is semi-locally-connected. This is at once apparent, since if L is such a continuum and $p \in L$, then for any ϵ and any $\delta < \epsilon$, at most a finite number of components of $M - V_\delta(p)$ can intersect $M - V_\epsilon(p)$. However, a non-compact locally connected set may fail to be semi-locally-connected, e. g., let L be the sum of the rays $\theta = 1/n$, $\rho \geq 0$, $(n = 1, 2, 3, \cdots)$.

(ii). Let A be the sum of the periphery of a rectangle R and an infinite sequence of altitudes of R converging to the left-hand vertical side of R. Let B be the sum of a triangle OPQ together with an infinite sequence of intervals OQ_i where Q_i is on PQ and the sequence $\{Q_i\}$ converges to Q. The continua A and B are semi-locally-connected but not locally connected.

(iii). Let S be the sum of a circle C and a spiral approaching C asymp-

[2] See *Mathematische Zeitschrift*, vol. 9 (1921), pp. 44-66.

totically. Let T be the sum of the interval $(-1, 1)$ of the y-axis and a portion of the curve $y = \sin 1/x$ approaching this interval asymptotically. The continua S and T are not semi-locally-connected. However, if we take the topological product of S or T by an interval or a circle, we obtain interesting examples of continua which are semi-locally-connected but not even arcwise connected. (See § 3 below).

(iv). Let C_0 be a circle in the plane with center at the origin and radius 1; for each $n > 0$, let C_n be a circle concentric with C_0 and of radius $1 + 3/(n+2)$. Let S_1 denote the spiral $(\rho - 1)\theta = 2\pi$, $(2\pi \leq \theta < \infty)$, S_2 the spiral $(\rho - 1)(\theta - 2\pi/3) = 2\pi$, $(2\pi + 2\pi/3 \leq \theta < \infty)$ and S_3 the spiral $(\rho - 1)(\theta - 4\pi/3) = 2\pi$, $(2\pi + 4\pi/3 \leq \theta < \infty)$. Finally, let $H = \sum_0^\infty C_i + \sum_1^3 S_i$. Then H is a semi-locally-connected plane compact continuum. However, it will be noted that H is not arcwise connected.

3. THEOREM. *In order that M be semi-locally connected at a point $p \in M$ it is necessary and sufficient that there exist arbitrarily small neighborhoods of p (i. e., open sets containing p) whose complements in M have only a finite number of components.*

Sufficiency. For $\epsilon > 0$, find a neighborhood $U \subset V_\epsilon(p)$ so that $M - U$ has only a finite number of components. Take δ so that $V_\delta(p) \subset U$. Then $M - V_\epsilon(p)$ is contained in the finite number of components of $M - V_\delta(p)$ which contain $M - U$.

Necessity. For $\epsilon > 0$ find δ so that $M - V_\epsilon(p)$ is contained in the sum N of a finite number of components of $M - V_\delta(p)$. Let $U = M - N$. Then U is open and $U \subset V_\epsilon(p)$.

Remark. If M is compact and semi-locally connected, it is readily seen that M is *uniformly semi-locally connected*, i. e., the δ in the definition can be chosen independent of the point p. This is easily established directly, or it can easily be deduced from the above theorem.

4. PRODUCT THEOREM. *If either one of the connected metric spaces A and B is semi-locally connected, so also is their topological (cartesian) product $A \times B$.*

Let us denote the space $A \times B$ by P. In proving this theorem we will suppose P metrized by the formula [3]

$$\rho(p_1, p_2) = \max[\rho(x_1, x_2), \rho(y_1, y_2)], \quad p_1 = (x_1, y_1), \quad p_2 = (x_2, y_2).$$

[3] See Kuratowski, *Topologie* I, p. 88.

The symbol $V_r(z)$ will be used to denote the r-neighborhood of the point z in whichever of the spaces A, B, or P, z is an element. We will suppose that A is semi-locally-connected.

Now take $p \in P$, $p = (a, b)$, $a \in A$, $b \in B$ and any $\epsilon > 0$. By the semi-local-connectivity of A, there exists a $\delta > 0$ such that $A - V_\epsilon(a) \subset A_1 + A_2 + \cdots + A_n$, where each A_i is a component $A - V_\delta(a)$. Let $P_i = A_i \times B$, $i = 1, 2, \cdots, n$. Then P_i is a continuum and $P_i \subset P - V_\delta(p)$. Now let $x = (a', b') \in P - V_\epsilon(p)$. Then either $a' \in A - V_\epsilon(a)$ or $b' \in B - V_\epsilon(b)$. In the first case, we have $x \in P_i$ for some i; and in the second case, $A \times b'$ is a continuum in $P - V_\epsilon(p)$ containing both $x = (a', b')$ and a point $y = (a_0, b')$ with $a_0 \in A - V_\epsilon(a)$. But then $y \in P_j$ for some j, so that x lies in the component Q_j of $P - V_\delta(p)$ containing P_j. Thus $P - V_\epsilon(p) \subset Q_1 + Q_2 + \cdots + Q_n$, and our theorem is proven.

Note: Let A be a non-locally-connected continuum, (e. g., the continuum S in § 2, (iii)) and let B be a semi-locally-connected continuum (e. g., an arc) and let $M = A \times B$. Let $T(M) = A$ be the "projection" of M into A, i. e., for each $x = (a, b) \in M$, $T(x) = a$. Then T is continuous, monotone and interior. However, M is semi-locally connected, A is not semi-locally-connected. Hence semi-local-connectedness is not invariant under such a transformation T.

5. Generalized *A*-sets. Let M be any metric space. A subset A of M will be called an *A-set* provided that $M - A = \Sigma G_i$ where G_i is open, $F(G_i)$ contains at most one point and $\delta(G_i) \to 0$, i. e., $M - A$ is the sum of a finite number or of a null sequence of open sets each having at most one boundary point.

THEOREM. *For any A-set $A \subset M$ and any connected set $Z \subset M$, $A \cdot Z$ is either vacuous or connected.*

Proof. Supposing the contrary, we have a separation

$$A \cdot Z = A_1 + A_2.$$

For any i, let $B_i^1 = B_i^2 = 0$ if $Z \cdot G_i = 0$; let $B_i^1 = Z \cdot G_i$ and $B_i^2 = 0$ if $Z \cdot G_i \neq 0$ and $F(G_i) \subset A_1$; and let $B_i^1 = 0$ and $B_i^2 = Z \cdot G_i$ if $Z \cdot G_i \neq 0$ and $F(G_i) \subset A_2$. Let

$$Z_1 = A_1 + \Sigma B_i^1, \qquad Z_2 = A_2 + \Sigma B_i^2.$$

Then clearly $Z = Z_1 + Z_2$ and $Z_1 \cdot Z_2 = 0$. If Z_1 contains a limit point x of Z_2, clearly $x \in A$, since each B_i^1 is open in Z. But since x is not a limit point of any single set B_i^2 and $\delta(B_i^2) \to 0$, there would exist a sequence of sets $B^2{}_{i_j} \to x$. This gives $F(B^2{}_{i_j}) \to x$, which is impossible since by definition $F(B^2{}_{i_j}) \subset A_2$.

II. Structure Development.

Throughout this part we shall suppose that M is a metric, connected, and semi-locally-connected space. In those cases where compactness is essential, it will be so specified in the statement of the result.

6. Complements of single points. Irreducible continua.

(6.1) *For any $x \in M$, each component R of $M - x$ is open.*

For let $p \in R$ and let $\epsilon = \frac{1}{2}\rho(x, p)$ and find a corresponding δ. Then if Q is the component of $M - V_\delta(p)$ containing p, we have $p \subset Q \subset R$; and since there are only a finite number of components of $M - V_\delta(p)$ containing points of $M - V_\epsilon(p)$, clearly p is interior to Q.

(6.11) *If M is compact, the components of $M - x$ form a null sequence at most.*

(6.2) *If $x \in M$ does not separate two points $a, b \in M$, there exists a closed and connected subset N of M with $a + b \subset N \subset M - x$.*

Proof. Let D be the component of $M - x$ containing a. Choose $\epsilon_1 = \frac{1}{2} \min [\rho(x, a), \rho(x, b)]$ and find δ_1. Take a sequence $\epsilon_1 > \delta_1 > \epsilon_2 > \delta_2 > \epsilon_3 \cdots > 0$. Let R_i be the component of $M - V_{\delta_i}(x)$ containing a and let $R = \Sigma R_i$. Since for any integer k there exists an integer K such that for $m, n > K$, $R_m \cdot [M - V_{\delta_k}(x)] = R_n \cdot [M - V_{\delta_k}(x)]$, it follows that R is closed in D. But since for each i, there are only a finite number of components of $M - V_{\delta_{i+1}}(x)$ containing points of $M - V_{\epsilon_{i+1}}(x)$ and since R_{i+1} is one of these, $R_{i+1} \supset R_i$, and $R_i \subset M - V_{\epsilon_{i+1}}(x)$, it follows that every point of R_i is interior to R_{i+1}. Hence R is open in M and thus certainly in D. Accordingly, since D is connected, $R = D$. Thus $R \supset b$. Hence $b \in R_k$ for some k and setting $N = \bar{R}_k$ we have $a + b \subset N \subset D \subset M - x$.

(6.21) *Each component of $M - x$ is strongly connected,* i. e., any two points of such a component R lie together in a closed and connected set contained wholly in R.

(6.22) *If x is a non-cut-point of M, (or if x cuts M into exactly k components), there exists an arbitrarily small neighborhood U of M such that $M - U$ is connected (or $M - U$ has exactly k components).*

This results at once from the proof of (6.2) together with § 3.

For any two points a and b of M, let us denote by $E(a, b)$ the set of all points x of M such that x separates a and b in M. From (6.2) we have at once

(6.23) *For any $a, b \in M$, $E(a,b) + a + b$ is closed.*

In case M is locally connected, every such set $E(a,b) + a + b$ is necessarily compact as well as closed,[4] whether M is compact or not. Of course, if M is compact and semi-locally-connected, it follows that every $E(a,b) + a + b$ is compact. However, if M is not compact, this does not follow. For let I be the interval $(0 \leq x < 1)$ of the x-axis, let I_i be the interval $(x = 1 - 1/i, 0 \leq y \leq 1)$ for $i = 1, 2, \cdots$, and let J be the interval $(y = 1, 1 \leq x \leq 2)$. Then if $M = I + J + \Sigma I_i$, $a = (0,0)$, $b = (2,1)$, it is readily seen that although M is semi-locally-connected, $E(a,b) + a + b$ is not compact.

In spite of this situation, it can (and will now) be shown that in case $M = E(a,b) + a + b$, this set is necessarily compact. In other words we will prove the following theorem:[5]

(6.3) *If M is an irreducible continuum between two of its points a and b, M is a simple arc from a to b.*

Proof. Since it follows at once from (6.2) that every point of $M - (a + b)$ must separate a and b in M and hence that $M = E(a,b) + a + b$, clearly all we have to prove is that M is compact.

If this is not so, there exists an infinite sequence of distinct points $x_1, x_2, \cdots$ of $E(a,b)$ having no limit point. Since[6] the points of $E(a,b) + a + b$ are arranged in a natural linear order from a to b, we may suppose without loss of generality that $[x_i]$ is a monotone increasing sequence in that order from a to b, i.e., x_i separates M into two connected separated sets: $M - x_i = A(x_i) + B(x_i)$ where $a + \sum_{1}^{i-1} x_j \subset A(x_i)$, $b + \sum_{i+1}^{\infty} x_j \subset B(x_i)$. Now since $A = \sum_{1}^{\infty} A(x_i)$ is open and contains a but does not contain b, it cannot be closed. Hence there exists a point $y \in M - A$ which is a limit point of A. Since y is not a limit point of $[x_i]$, it follows by (6.1) and (6.2) that

---

[4] See my papers in *Bulletin of the American Mathematical Society*, vol. 33 (1927), p. 685; *Transactions of the American Mathematical Society*, vol. 32 (1930), p. 927.

[5] Compare with Hausdorff, *Mengenlehre* 1927. p. 222. That Hausdorff's assumption of compactness is superfluous has been pointed out by the author elsewhere.

It is of interest to note that it follows from (6.3) that assuming every subcontinuum of a compact continuum M semi-locally-connected is equivalent to assuming every subcontinuum of M locally connected. For it would follow by (6.3) that every irreducible continuum in M between two points is a simple arc; and by a theorem of Kline [*Fundamenta Mathematicae*, vol. 10 (1927), p. 298], it follows that every subcontinuum of M is locally connected.

[6] See my paper in *Transactions of the American Mathematical Society*, vol. 33 (1931), pp. 444-454.

there exists a neighborhood U of y such that the component R of $M - U$ containing a contains $\sum_1^\infty x_i$. But since R is connected and thus so is $R + B(x_i)$ for each i, clearly this gives $A \subset R$, contrary to the fact that y is a limit point of A. Thus M is compact and our theorem is proven.

7. Cyclic elements. By a 0-th order cyclic element or E_0-set of M is meant either a cut point of M or a sub-continuum of M (possibly degenerate) which is maximal with respect to the property of having no cut point.[7] If p is a non-cut point of M, the set M_p consisting of p together with all points of M which cannot be separated in M from p by any single point is called[8] a *simple link* of M. Since it is known[9] that, for any compact continuum M, any point of M is either a cut-point, an end point (i. e., point of Menger-Urysohn order 1) or a point of exactly one non-degenerate simple link of M, since it will be shown below in this section that, for M compact and semi-locally-connected, the non-degenerate E_0-sets and simple links are identical; and since we make no use of cyclic elements of higher order than the E_0's, in the sections which are to follow these sets will be referred to simply as the *cyclic elements* of M.

(7.1) *If M is compact,*[10] *no simple link of M can have a cut point. Thus the simple links and the E_0-sets are identical.*

Proof. Let E be a simple link defined by the non-cut point p of M. Suppose, contrary to our theorem, that E has a cut point x. Let $E - x = E_a + E_b$ be a separation, $a \in E_a$, $b \in E_b$. Regardless of the relation between p and x, x cannot separate a and b in M. Hence by (6.2) there exists a continuum N with $a + b \subset N \subset M - x$ and we may suppose N is irreducible from a to b. But clearly $N \subset E$; for if not, some point $q \in N$ gives a separation $M - q = M_1 + M_2$ where $a + b + p \subset M_1 + q$ and $M_2 \cdot N \neq 0$; and this is impossible since $N \cdot (M_1 + q)$ is a proper subcontinuum of N containing both a and b. Thus $N \subset E$, which is impossible since N does not contain x.

(7.2) *If M is compact, if E is any E_0-set in M and R is any component of $M - E$, there exists a point x of E such that R is a component of $M - x$.*

Proof. Let $a \in R$ and let b be a non-cut point of M belonging to E. By

[7] See my paper in *American Journal of Mathematics*, vol. 56 (1934), pp. 133-146.

[8] See R. L. Moore, *Foundations of Point Set Theory*, p. 72.

[9] See a forthcoming paper by J. L. Kelly, abstracted in the *Bulletin of the American Mathematical Society*, vol. 45 (1939), p. 74.

[10] It should be noted that throughout this part we are assuming M semi-locally-connected.

(6.23), $E(a,b)+a+b$ is closed and hence compact. Since $E(a,b)+a \subset \bar{R}$, either there exists a last point x of $E(a,b)$ in the order (a,b) or else b is a limit point of $E(a,b)$ and hence of R. But since b is a non-cut point of M, for any $c \in E-b$ there exists a continuum N with $a+c \subset N \subset M-b$; and since $N \supset E(a,b)$, b cannot be a limit point of $E(a,b)$. Thus there exists a last point x in $E(a,b)$. Now we must have $x \in E$. For if not, there is a point y in $E(x,b)$ and clearly $y \in E(a,b)$. Thus $x \in E$; whence, $R \subset M-x$; and since x separates R from $E-x$, it follows that R is a component of $M-x$.

8. *A* sets. Throughout this section it will be assumed that M is compact as well as semi-locally-connected.

(8.1) *If M is compact, the following three properties for subcontinuum A of M are equivalent:*

(a) if A contains more than one point of a cyclic element E_0 of M, we have $A \supset E_0$.

(b) for any $p \in A$ we have either $p = A$ or $p \subset$ some E_0-set $\subset A$.

(c) the boundary of each component of $M-A$ is a single point.

Proof. (*a*) *implies* (*b*). For let $p \in A$. If p is either a cut point or an end-point the result is obvious. If not, there exists[9] a unique, non-degenerate E_0 containing p. If $E_0 \cdot A \neq p$, we have $E_0 \subset A$ by (a). Hence we may suppose $E_0 \cdot A = p$. Let $x \in E_0 - p$, $y \in A - p$. Since p is not a cut point, there exists a continuum N with $x+y \subset N \subset M-p$. The component Q of $N - N \cdot A$ containing x has a limit point q in A, $q \neq p$. Since $A \supset p+q \subset E_0+Q+q$ and $A \cdot (E_0+Q+q) = p+q$, clearly p and q are not separated in M by any single point. Whence, $p+q \subset E_0 \subset A$ by (a).

(b) *implies* (a). For let $E_0 \cdot A \supset p+q$. Let N be an irreducible continuum in A from p to q. Clearly $N \subset E_0$. Hence $E_0 \cdot A$ contains a non-cut point x of M. Whence $x \subset E_0 \subset A$.

(a) *implies* (c). For suppose the boundary of a component R of $M-A$ contains two distinct points p and q. Then since $(R+p+q) \cdot A = p+q$, there exists an E_0 with $p+q \subset E_0 \subset A$. Whence, $R \subset M-E_0$. But this is impossible since the component of $M-E_0$ containing R can have only one boundary point.

(c) *implies* (a). For let $E_0 \cdot A \supset p+q$. Suppose there exists $x \in E_0 - E_0 \cdot A$. Let R be the component of $M-A$ containing x and let $y = \bar{R} - R$. Since $E_0 - y$ is connected and contains x and at least one of the points p and q, say p, there is a continuum N with $p+x \subset N \subset M-y$. But clearly N contains a point of $\bar{R}-R$ contrary to $y \cdot N = 0$.

(8.2) *If R is a component of the complement of a continuum A in M satisfying* (a), (b), *or* (c), *and $y = \bar{R} - R$, then R is a component of $M - y$.*

For let D be the component of $M - y$ containing R. If $D \cdot A \neq 0$, we can find a continuum N in D intersecting both R and A; and this is impossible since clearly $N \cdot A$ would contain a limit point of R, contrary to $\bar{R} - R = y$. Hence $D \cdot A = 0$, which gives $D = R$.

(8.21) *Every component R of $M - A$ is open in M and strongly connected.*

(8.3) *Any subcontinuum A of M satisfying* (a), (b) *or* (c) *of* (8.1) *is an A-set in M.*

Proof. Since every component of $M - A$ is open and has only one boundary point, we have only to show that the components of $M - A$ form at most a null sequence. Suppose, on the contrary, that there exists a sequence of components $R_1, R_2, \cdots$ of $M - A$ converging to a non-degenerate limit continuum K. By (8.21) we have $K \subset A$. Now if $p_i = F(R_i)$, by (8.2) it follows that there are infinitely many of the p_i's. Hence we may suppose $p_i \to p \in A$. But it readily follows that M cannot be semi-locally-connected at p.

(8.31) *If Z is connected, $A \cdot Z$ is either connected or vacuous.*

(8.32) *The product of any number whatever of A-sets is itself an A-set.*

This follows at once from property (b) of (8.1).

For the remainder of this section, A will designate any A-set in M.

(8.4) *If Z is a* $\left\{\begin{matrix} \text{locally connected} \\ \text{semi-locally-connected} \end{matrix}\right\}$ *subset of M, so also is $A \cdot Z$ when non-vacuous.*

The first result is an obvious consequence of (8.31).

To prove the second one, let Z be semi-locally connected and let $x \in A \cdot Z$. For any $\epsilon > 0$, let δ be found so that $Z - Z \cdot V_\epsilon(x)$ is contained in the sum of a finite number of components $Z_1, Z_2, \cdots, Z_n$ of $Z - Z \cdot V_\delta(x)$. Then, since by (8.31) each of the sets $A \cdot Z_i$ is connected and since $\sum_1^n A \cdot Z_i \supset A \cdot Z - A \cdot Z \cdot V_\epsilon(x)$, it follows that $A \cdot Z$ is semi-locally-connected.

(8.5) *If two points $x, y \in A$ are separated in A by a subset X of A, then x and y are also separated in M by X.*

For let $A - X = A_x + A_y$ be a separation. Let $M_x = A_x +$ all components R of $M - A$ with $F(R) \subset A_x$ and let $M_y = A_y +$ (all other com-

ponents of $M - A) = M - X - M_x$. Then $M - X = M_x + M_y$ is a separation between x and y as is seen by a proof almost identical with that given in § 5.

Essentially the same proof suffices to establish

(8.6) *If* $A = H + K$ *is any division of* A *into* $\left\{\begin{matrix} sets \\ closed\ sets \\ continua \end{matrix}\right\}$ H *and* K, *there exists a division* $M = L + N$ *into* $\left\{\begin{matrix} sets \\ closed\ sets \\ continua \end{matrix}\right\}$ L *and* N *such that* $L \cdot N = H \cdot K$, $L \supset H$, $N \supset K$.

As an immediate consequence of the "continua" part of (8.6) we have

(8.7) *If* M *is a compact, semi-locally-connected and unicoherent continuum, so also is every cyclic element of* M_0.

Thus unicoherence is "E_0-reducible" for semi-locally-connected continua. It is well known that this is true for locally connected continua. That it is not true for continua in general is shown by the continuum S in example (iii) of § 2. On the other hand, unicoherence is E_0-extensible for arbitrary compact continua.

9. Continua of convergence.

THEOREM. *If* $[K_n]$ *is an infinite sequence of disjoint continua in* M *converging to a continuum* K, *there exists a cyclic element* E *containing* K *and such that* $K = \lim (E \cdot K_n)$.

Proof. As any two points of K are conjugate, clearly there exists a cyclic element E containing K. Also, $\limsup (EK_n) \subset K$. We have to show $\liminf (E \cdot K_n) \supset K$. Suppose, on the contrary, that there is a point x of K not belonging to $\liminf (E \cdot K_n)$. Then there exists an $\epsilon > 0$ and an infinite sequence of integers n_i such that $K_{n_i} \cdot V_\epsilon(x) \neq 0 = E \cdot K_{n_i} \cdot V_{3\epsilon}(x)$, for each i. Now for each i, $K_{n_i} \cdot V_{3\epsilon}(x)$ contains a continuum N_i of diameter $> \epsilon$. Since $N_i \subset M - E$, it follows by (8.3) that there exists a component R of $M - E$ containing infinitely many of the sets N_i. But clearly this is impossible, since $F(R)$ is a single point, whereas $\limsup N_i \subset E \subset M - R$ and $\delta(\limsup N_i) > \epsilon$.

(9.1) *The non-degenerate cyclic elements form a null sequence at most.*

For otherwise a sequence of disjoint cyclic elements $[E_i]$ could be found, with $\delta(E_i) > d > 0$ $(i = 1, 2, \cdots)$, which converges to a continuum K. The

above theorem then would give a cyclic element E so that $E_i \cdot E$ is non-degenerate for almost all i, which clearly is impossible.

(9.2) *Any continuum of convergence of M is a continuum of convergence of some cyclic element of M.*

10. Cyclic chains. For any two points $a, b \in M$, the product of all the A-sets in M containing $a + b$ will be called the *cyclic chain* from a to b and is denoted by $C(a, b)$.

THEOREM. *$C(a, b) = E(a, b) + a + b +$ all cyclic elements E such that $E \cdot [E(a, b) + a + b]$ consists of exactly two points.*

For let X denote the set on the right. Then X is an A-set. For since $E(a, b) + a + b$ is compact, X is compact. Also X is connected. For if not we have a separation $X = X_a + X_b$. Since $E(a, b) + a + b = K$ is linearly ordered and compact, there is a last point x of K in X_a and a first point y of K in X_b following x. Now x and y must be conjugate; for if z separates x and y it readily follows that $z \in E(a, b)$, contrary to the fact that there is no point of $E(a, b)$ "between" x and y. Hence there exists an E_0, say N, with $x + y \subset N$. But since clearly N cannot contain three points of K, we have $N \cdot K = x + y$ so that $N \subset X$ contrary to $x \in X_a$, $y \in X_b$. Thus X is connected. Now for any $x \in X$, there exists an E_0 with $x \subset E_0 \subset X$. This follows by the definition of X except when $x = a$ or $x = b$. Suppose $x = a$. Now if a is a cut point or an end-point of M, again our result follows. If not, let E_0 be the (non-degenerate) cyclic element containing a. Let R be the component of $M - E_0$ containing b and let $z = F(R)$. Then since $a \neq z$, we have $z \in E(a, b) + b$. Consequently $a + z = E_0 \cdot K$ and $E_0 \subset X$. A similar proof works for the case $x = b$. Thus X is an A-set.

Now since any A-set containing $a + b$ clearly must contain X, we have $C(a, b) \subset X$; and since X is an A-set containing $a + b$, we have $X \supset C(a, b)$. Whence $X = C(a, b)$.

11. IMBEDDING THEOREM. *Any compact semi-locally connected continuum M is topologically contained in a locally connected continuum L such that for every cyclic element C of L, $M \cdot C$ is a cyclic element of M. Furthermore, L can be chosen in the Hilbert parallelotope Q_ω such that every true cyclic element of C is a topological sphere whose interior contains no point of M.*

This theorem can be proved readily by well-known methods of reasoning from the following facts.

(i) *M has a cyclic chain development*, i. e., there exist in M two sequences

of distinct points $[p_i]$ and $[q_i]$ $(i = 1, 2, \cdots)$ such that the cyclic chains $C(p_i, q_i)$ have the following properties.

(a) $C(p_i, q_i) \cdot \sum_{j=1}^{i-1} C(p_j, q_j) = q_i \ (i = 1, 2, \cdots)$;

(b) setting $H_n = \sum_{i=1}^{n} C(p_i, q_i)$ and $H = \sum_{1}^{\infty} H_n$, every point of $M - H$ is an end point;

(c) the diameter of the components of $M - H_n$ approach 0 with $1/n$;

(d) $\lim C(p_n, q_n) = 0$.

This can be proved with the aid of the preceding sections by essentially the same argument used to prove it in the locally connected case.[11]

(ii) *Any separable metric space can be topologically imbedded in the unit sphere* $\Sigma x_i^2 = 1$ *of a Hilbert space or of* Q_ω.

To see this let H be the Hilbert space with points $(x_1, x_2, \cdots)$ where Σx_i^2 is convergent and let H_1 be the plane $x_1 = 1$ of H. Then the function

$$f(x_1, x_2, \cdots) = (1, \tfrac{1}{2}x_1, \tfrac{2}{3}x_2, \tfrac{3}{4}x_3, \cdots)$$

maps H topologically into H_1, as is readily verified. Furthermore, if for each $x = (1, x_2, x_3, \cdots) \in H_1$ we define

$$g(x) = (x_1/\sigma, x_2/\sigma, x_3/\sigma, \cdots) \text{ where } \sigma = \sum_{1}^{\infty} x_i^2, \ (x_1 = 1),$$

then g maps H_1 topologically into a subset of the unit sphere S: $\Sigma x_i^2 = 1$.

Hence $gf(H) \subset S$ is a topological mapping of H into a subset of S and of course gf maps Q_ω into a subset of $Q_\omega \cdot S$. Thus (ii) follows from the well-known theorem of Urysohn that any separable metric space can be topologically imbedded in H.

(iii) *In case* M *is a cyclic chain* $C(p, q)$, the theorem is proved as follows. Since the set $E(p, q) + p + q = K$ is compact and naturally ordered, there exists a topological, order preserving transformation $f(x)$ mapping this set into a subset of a linear interval ab of Q_ω so that $f(p) = a$, $f(q) = b$. Now by § 10, each true cyclic element E_i of M contains exactly two points x_i and y_i of $E(p, q) + p + q$. For each such E_i, let S_i be the sphere in Q_ω having the interval $f(x_i)f(y_i)$ of ab as a diameter and, using (ii), let us extend f to E_i in such a way that $f(E_i) \subset S_i$ and f is topological. Then clearly the extended f maps M topologically into a subset of $L = f(K) + \Sigma S_i$; and this latter set

[11] See Kuratowski and Whyburn, *Fundamenta Mathematicae*, vol. 16 (1930), pp. 305-331.

satisfies all the requirements of our theorem, since its cyclic elements are the S_i together with the individual points of $f(K)$.

COROLLARY. *If M is compact and semi-locally-connected and if for some m, each cyclic element of M is homeomorphic with a subset of the unit sphere in R^m, then M can be topologically imbedded in a compact, $(m-1)$-dimensional "cactoid" L in R^m which is the boundary of its unbounded complementary domain in R^m so that the intersection of M and each cyclic element of L is a cyclic element of M.*

This corollary would apply in case the dimensionality of M is finite. It is of interest to note that it may be regarded as an extension of the theorem of W. L. Ayres[12] that any locally-connected continuum every true cyclic element of which is a simple closed curve is homeomorphic with the boundary of a plane domain, since it reduces to that theorem in the case $m = 2$.

12. Cyclicly extensible and reducible properties. Convexity. The dominant feature of the preceding development of the structure of a compact semi-locally-connected continuum—and especially of the imbedding theorem—is the close analogy with the locally connected case. It is therefore to be expected that a large number of properties which are cyclicly extensible and reducible for locally connected continua would remain so for semi-locally-connected continua. We list here a few such properties without proofs—the proofs being not essentially different here from those in the locally connected case. (In some instances the conclusions have already been established above or they follow directly from the imbedding theorem.) We recall that a property P is cyclicly extensible provided "*each cyclic element of M has P*" implies "*M has P*" and is cyclicly reducible if the reverse implication holds.

The following properties are cyclicly extensible and reducible for compact semi-locally-connected continua M:

(i) *unicoherence.*
(ii) *local connectedness.*
(iii) *arcwise connectedness.*
(iv) *being a rational curve.*
(v) *having dimensionality* $\leq n$ $(n > 0)$.
(vi) *being topologically convex.*

In connection with property (vi), the following may be of interest. To say that a space X is topologically convex means that X is homeomorphic with a metrically convex space, or, in other words, that X has a convex metric, i. e.,

[12] *Fundamenta Mathematicae*, vol. 14 (1929), pp. 92-95.

a metric $\rho(x, y)$ such that for any $a, b \in X$ and any t with $0 \leq t \leq \rho(a, b)$ there is a single $c \in X$ such that $\rho(a, c) = t$ and $\rho(a, c) + \rho(c, b) = \rho(a, b)$. Now if each cyclic element of M has this property, clearly each such element is locally connected. Hence, by (ii), M is locally connected. Thus the extensibility of (vi) follows from (ii) and the previously known [11] locally connected case. (The same situation would obtain in case of the property of being an absolute retract.[13]) Now in the imbedding theorem, we can add to L the interior of each of its spherical cyclic elements and thus obtain a continuum L^* with convex cyclic elements without losing any of the asserted relationship between L and M. Hence L^* is topologically convex and thus we have the

THEOREM. *Any compact, semi-locally-connected continuum M has a metric which is convex relative to the cyclic elements of M, i. e., a metric such that for any $a, b \in M$ and any $x \in E(a, b)$, $\rho(a, x) + \rho(x, b) = \rho(a, b)$.*

13. Retracts. By virtue of the imbedding theorem, the study of compact semi-locally-connected continua is equivalent to the study of particular kinds of subcontinua of local connected continua. Furthermore, the containing continua may be taken so that they are absolute retracts. Hence the following theorem becomes significant.

THEOREM. *In order that a subcontinuum N of a compact semi-locally-connected continuum M be a retract of M it is necessary and sufficient that $N \cdot E_0$ be a retract of E_0, for every cyclic element E_0 of M.*

Proof. The condition is necessary. For let N be a retract of M and let E be any E_0-set in M. Let $f(x)$ be a function retracting M into N, i. e.,

$$f(x) \subset N \text{ for } x \in M$$
$$f(x) = x \text{ for } x \in N.$$

Let us define

$g(x) = x$ for $x \in E$
$g(x) = F(C_x)$ if x non $\in E$, where C_x is the component of $M - E$ containing x.

Then g retracts M into E. Furthermore gf retracts M (hence also E) into $N \cdot E$.

The condition is sufficient. For let $N \cdot E_0$ be a retract of E_0 for every E_0 in M. Since any A-set in M is a retract of M, we can retract M into the smallest A-set A in M containing N by a monotone retracting function $r(x)$. Now if we show that A can be retracted into N by a function $f(x)$, clearly fr

[13] See Borsuk, *Fundamenta Mathematicae*, vol. 18 (1932), pp. 198-213.

will retract M into N. Hence we may suppose without loss of generality that $M = A$; in other words, we have $N \cdot E_0 \neq 0$ for each E_0 in M and $N \cdot E_0$ is non-degenerate when E_0 is non-degenerate. Let $E_1, E_2, \cdots$ be the non-degenerate E_0-sets in M. For each i, let $f_i(x)$ be a function retracting E_i into $N \cdot E_i$. Let us define

$$f(x) = x \text{ for } x \in N,$$
$$f(x) = f_i(x) \text{ for } x \in E_i.$$

Clearly $f(x)$ is uniquely defined. It is also continuous. For if x non $\in N$, there exists an i and an open set U in M such that $x \subset U \subset E_i$ and on U, $f \equiv f_i$; if $x \in N$, and $\epsilon > 0$, let $E_{i_1}, E_{i_2}, \cdots, E_{i_n}$ be the E_0 sets which contain x and are of diameter $> \epsilon$. Since the functions f_i are continuous and $f_i(x) = x$, for any $\epsilon > 0$ a $\delta > 0$ exists such that if $y \in E_{i_j} \cdot V_\delta(x)$, we have $f_{i_j}(y) = f(y) \subset V_\epsilon(x)$. There exists a $d > 0$ such that if $i \neq i_j$, $(j = 1, \cdots, n)$ and $E_i \cdot V_d(x) \neq 0$ we have $E_i \subset V_\epsilon(x)$. Thus if $y \in E_i \cdot V_d(x)$ we have $f_i(y) = f(y) \subset E_i \subset V_\epsilon(x)$. Finally if y non $\in E_i$ for all i, we have $y \in N$ and $f(y) = y$. Hence f is continuous and retracts M into N.

III. Plane Continua.

We proceed at once to establish the theorem mentioned in the introduction. It will be noted that except for (6.22), which is readily established directly from the definition of semi-local-connectedness, the proof to be given is entirely *independent* of the preceding results in this paper.

14. THEOREM. *If M is a compact semi-locally-connected plane continuum, the boundary of each complementary domain of M is locally connected.*

Proof. Let R be a complementary domain of M. Clearly we may suppose R bounded, since an inversion of the plane can be used to throw R into a bounded region complementary to the image of M. We shall show that R has property S, i. e., R is the sum of a finite number of arbitrarily small connected sets.

If this is not so, there exists a sequence of points $p_1, p_2, \cdots$ of R converging to a point $p \in M$ and an $\epsilon > 0$ such that for no i, j can p_i and p_j be joined by a connected subset of R of diameter $< 4\epsilon$.

Let us suppose first that p is not a cut point of M. Then by (6.22) there exists a $\delta > 0$ and a component N of $M - M \cdot V_\delta(p)$ which contains $M - M \cdot V_\epsilon(p)$. Let C_δ and C_ϵ denote the circles $F[V_\delta(p)]$ and $F[V_\epsilon(p)]$ respectively. Now since almost all the points p_i lie in $V_\delta(p)$, we may suppose that all of them do. Let $p_1 p_2$ be an arc in R from p_1 to p_2.

There must exist a subarc xy of $p_1 p_2$ such that $x + y \subset C_\delta$, $xy - (x + y)$

lies without C_δ and x and y cannot be joined by a continuum of R lying in $V_{2\epsilon}(p)$. For if p_1p_2 contained no such arc, we could find a finite number of subarcs $x_1y_1, x_2y_2, \cdots, x_ny_n$ of p_1p_2 such that $x_i + y_i \subset C_\delta$, $x_iy_i - (x_i + y_i)$ lies without C_δ and $p_1p_2 - \Sigma x_iy_i \subset V_\epsilon(p)$. Furthermore, for each i, there would exist a continuum H_i in $R \cdot V_{2\epsilon}(p)$ containing $x_i + y_i$. Thus $(p_1p_2 - \Sigma x_iy_i) + \Sigma H_i$ would be a continuum in $R \cdot V_{2\epsilon}(p)$ containing both p_1 and p_2. This is impossible by the choice of the points $p_1, p_2, \cdots$.

Let u and v be points separating x and y on C_δ. Now the continuum N can intersect only one of the arcs xuy, xvy of C_δ, since neither N nor xy contains a point of $V_\delta(p)$ and $N \cdot xy = 0$. Hence xuy, say, contains no point of N. Let J denote the simple closed curve $xy + xuy$ and let I denote that region complementary to J which does not contain N. Let 0 be a point of $I \cdot V_\epsilon(p)$ which can be joined to u by an arc ou such that $ou - o$ lies in $I \cdot V_\epsilon(p)$. Then by a theorem of R. L. Moore,[14] there exists a simple closed curve $C \subset J + C_\epsilon$ such that if E is the complementary region of C containing 0, then $E \subset I \cdot V_\epsilon(p)$. But now if x_1 and y_1 denote the first and last points respectively of C_ϵ on xy in the order x, y, it follows that since $ou - o \subset E$, the arc x_1xuyy_1 of J is contained in C. Then if x_1wy_1 is the other arc of C from x_1 to y_1 we have $x_1wy_1 \subset C_\epsilon + xy \cdot V_\epsilon(p) \subset V_{2\epsilon}(p)$. But since $x_1wy_1 \subset \bar{I}$ and $\bar{I} \cdot N = 0$, we have $x_1wy_1 \cdot N = 0$; and since $C_\epsilon \cdot M = C_\epsilon \cdot N$ and $xy \cdot M = 0$, this gives $x_1wy_1 \cdot M = 0$. Whence, $x_1wy_1 \subset R \cdot V_{3\epsilon}(p)$. Thus the arc xx_1wy_1y of C is contained in $R \cdot V_{2\epsilon}(p)$, which is impossible by the choice of the arc xy. Hence we obtain a contradiction if p is a non-cut point of M.

Now in general, let us choose a point q outside $V_\epsilon(p)$ and join p_i to q by an arc p_iq in R. Let q_i be the first point of C_ϵ on p_iq. The sequence of arcs p_iq_i contains a subsequence (which we may suppose is the whole sequence) converging to a non-degenerate continuum $H \subset M$. Since every point of H within $V_\epsilon(p)$ is, like p, the limit of a sequence of points of R no pair of which can be joined in R by a connected set of diameter $< 3\epsilon$, it follows by the above argument that every point of H is a cut point of M. But then [6] there exists a point p' of $H \cdot V_\epsilon(p)$ which is of order 2 in M. Hence there exists a simple closed curve K in $V_\epsilon(p)$ such that $K \cdot M = c + d$, $c, d \in M$, and the complementary region D of K containing p' lies in $V_\epsilon(p)$. But there exists points p'_i, p'_j, p'_k from distinct arcs p_iq_i, p_jq_j, p_kq_k which lie in D. Since each of these arcs intersects $K - (c + d)$, two of them, say p_iq_i, p_jq_j must intersect one of the two components, say F, of $K - (c + d)$. Then $p_iq_i + p_jq_j + F$ is a connected subset of R of diameter $< 3\epsilon$ containing both p_i and p_j. Thus in any case we get a contradiction if we assume R does not have property S.

[14] *Transactions of the American Mathematical Society*, vol. 17 (1916), pp. 131-164.

Thus R has property S and hence by a theorem of R. L. Moore's [15] the boundary of R is locally connected.

COROLLARY 1. *If M is a compact plane semi-locally-connected cyclic continuum, the boundary of every complementary domain of M is a simple closed curve.*[16]

COROLLARY 2. *In order that a compact plane semi-locally-connected continuum be locally connected it is necessary and sufficient that the complementary domains form a null sequence at most.*

COROLLARY 3. *If M is a compact plane semi-locally-connected continuum whose first Betti number $p^1(M)$ is finite, then M is locally connected.*

15. In conclusion it may be remarked that the property of semi-local-connectedness clearly may be formulated in combinatorial terms; and in this form the definition can be enlarged so as to cover higher dimensional connectedness. Also it is possible to define combinatorially properties of domains of higher dimensionalities which may be regarded as higher dimensional analogues of property S. In view of the proof of the theorem in § 14 it seems probable that some sort of dual relationship could be found between the higher dimensional semi-local-connectedness properties of a compact set in a Euclidean space and the S-properties of its complementary domains. It is hoped that these considerations will be further developed in a later paper.

UNIVERSITY OF VIRGINIA.

---

[15] R. L. Moore, *Fundamenta Mathematicae*, vol. 3 (1922), pp. 232-237.

[16] Since, as was pointed out to me by R. L. Wilder after this paper was submitted for publication, it follows by the argument given under (6.2) that in the case of a cyclic continuum, "semi-local-connectedness" is equivalent to Wilder's property "local 0-avoidability" it results that this corollary is equivalent to the first corollary following Theorem 8a on p. 196 of his paper "Sets which satisfy certain avoidability conditions" cited in reference 1.

CONCERNING LOCAL CONNECTEDNESS UNDER THE INVERSE OF CERTAIN CONTINUOUS TRANSFORMATIONS.*

By W. T. PUCKETT, JR.

(1) This paper concerns itself with conditions under which local connectedness is preserved under the inverse of certain continuous transformations $T(A) = B$, defined on a compact metric space A. Interior transformations[1] are studied in particular, and the principal result here is: In order that local connectedness be invariant under the inverse of an interior transformation $T(A) = B$, it is necessary and sufficient that the inverse of every point b of B be locally connected relative to A. A continuous transformation satisfying the above condition is said to have property β relative to (A, B).

In the latter part of the paper it is assumed that $T(A) = B$ is a light transformation.[2] From a general theorem established, the following results are obtained: If the inverse of each point b of B is finite (countable), then the property of being a regular (rational) curve is invariant under the inverse of T. In case T is interior and finite, local connectedness is invariant under the inverse.

(2) A continuous transformation $T(A) = B$ is *interior*[1] provided every open set in A maps into an open set in B. Eilenberg[3] has shown that $T(A) = B$ is interior on a compact set A if and only if the convergence of any sequence of points $\{b_n\}$ to a point b in B implies the convergence of $\{T^{-1}(b_n)\}$ to $T^{-1}(b)$ in A.

A continuous transformation $T(A) = B$ has *property β relative to* (A, B) provided that for every point b of B, the set $T^{-1}(b)$ is locally connected relative to A, i. e., for every $\epsilon > 0$ there exists a $\delta > 0$ such that if a and a' belong to the inverse of some point b of B and $\rho(a, a') < \delta$, then a and a' lie together in a connected subset of A of diameter less than or equal to ϵ. An immediate

* Received July 29, 1938; Revised December, 1938; Presented to the American Mathematical Society, April 15, 1938.

[1] See Stoilow, *Annales Scientifiques des l'École Normale Supérieure*, vol. 63 (1928), pp. 347-382. We follow the custom of most later writers in omitting Stoilow's second condition from the definition.

[2] See Whyburn, "Interior transformations on compact sets," *Duke Mathematical Journal*, vol. 3 (1937), p. 370.

[3] See "Sur les transformations d'espaces métriques en circonférences," *Fundamenta Mathematicae*, vol. 24 (1935), p. 174.

consequence of the definition is: *If A is a locally connected compact set, then every continuous transformation $T(A) = B$ has property β relative to (A, B).*

(2.1) *Let $T(A) = B$ be an interior transformation of the compact set A having property β relative to (A, B). If p is a point of A such that B is locally connected at $T(p)$, then A is locally connected at p.*

It will be shown that for any $\epsilon > 0$ there exists a neighborhood V of p such that if q is a point of V, then q and p lie together in a connected subset of $U(p, \epsilon)$.[4] Let δ be determined by using $\epsilon/3$ for ϵ in the definition of property β. Since T is interior and B is locally connected at $T(p)$, there exists a neighborhood V of p such that $T(V)$ is contained in a connected subset of $T[U(p, \delta/4)]$.

In order to show that V has the desired property it suffices to show that for any $d > 0$, any point q of V can be d-chained to p in $U(p, \epsilon)$, i. e., there exist points $q = a_0, a_1, a_2, \cdots, a_n = p$ contained in $U(p, \epsilon)$ with $\rho(a_i, a_{i+1}) < d$. Since T is interior, there exists[5] a σ such that for every point a of A the neighborhood $U[T(a), \sigma]$ is contained in $T[U(a, d)]$. Since $T(V)$ is contained in a connected subset of $T[U(p, \delta/4)]$ there exists a σ-chain, $T(q) = b_0, b_1, b_2, \cdots, b_k = T(p)$, in $T[U(p, \delta/4)]$ and, consequently, there exist points a'_i of $U(p, \delta/4)$ such that $T(a'_i) = b_i$. From the selection of σ it follows that $U(q, d)$ must contain a point a_1 such that $T(a_1) = b_1$. It may be assumed that $d < \delta/4$. Hence a_1 must belong to $U(p, \delta/2)$, since q belongs to $U(p, \delta/4)$. Therefore $\rho(a'_1, a_1) < \delta$ and, consequently, there exists a connected set K_1 containing a'_1 and a_1 with $\delta(K_1)$[6] $< \epsilon/3$, because of property β. Therefore $q + a_1 + K_1$ is contained in $U(p, 2/3\epsilon)$ and contains a d-chain from q to a'_1. Now $U(a'_1, d)$ contains a point a_2 lying in $U(p, \delta/2)$ such that $T(a_2) = b_2$. Therefore, there exists a connected set K_2 with $\delta(K_2) < \epsilon/3$ containing a_2 and a'_2. Hence $q + a_1 + K_1 + a_2 + K_2$ contains a d-chain from q to a'_2 lying in $U(p, 2/3\epsilon)$. By continuing this process one arrives at a connected set K_k of diameter less than or equal to $\epsilon/3$, which contains p and a point a_k of $U(p, \delta/2)$ such that $T(a_k) = b_k$. Hence $K = q + \sum_{i=1}^{k} (a_i + K_i)$ is contained in $U(p, 2/3\epsilon)$ and contains a d-chain from q to p, and the theorem is demonstrated.

The definition of property β and the above theorem yield at once the following theorem:

---

[4] The symbol $U(p, \epsilon)$ denotes the ϵ-neighborhood of p, i. e., the set of all points x of A such that $\rho(x, p) < \epsilon$.

[5] See Whyburn, "The mapping of Betti groups under interior transformations," *Duke Mathematical Journal*, vol. 4 (1938), p. 1.

[6] The diameter of a set K is denoted by $\delta(K)$.

(2.11) *Let $T(A) = B$ be an interior transformation of the compact set A, and let B be locally connected. In order that A be locally connected it is necessary and sufficient that T have property β relative to (A, B).*

(2.2) *Let $T(A) = T_2T_1(A) = B$, where $T_1(A) = A'$ and $T_2(A') = B$ are continuous. If T has property β relative to (A, B), then T_1 has property β relative to (A, A').*

For an arbitrary $\epsilon > 0$ let $\delta > 0$ be determined by property β of T relative to (A, B). Let a' be any point of A', and let a_1 and a_2 be any two points of $T^{-1}(a')$ such that $\rho(a_1, a_2) < \delta$. Since a_1 and a_2 are points of $T^{-1}[T_2(a')]$ they lie together in a connected subset of A of diameter less than or equal to ϵ. Hence T_1 has property β relative to (A, A').

(3) An application of the principal theorem of the preceding section is given by the 0-regular transformation defined by A. D. Wallace.[7] The continuous transformation $T(A) = B$ is *0-regular* provided the convergence of the sequence of points $\{b_n\}$ to the point b in B implies the 0-regular convergence[8] of the sets $\{T^{-1}(b_n)\}$ to $T^{-1}(b)$, i. e., for each $\epsilon > 0$ there exist positive numbers δ and N such that if $n > N$, any pair of points x and y of $T^{-1}(b_n)$ with $\rho(x, y) < \delta$ lie together in a continuum in $T^{-1}(b_n)$ of diameter less than ϵ. That a 0-regular transformation on a compact set is interior follows at once from the Eilenberg characterization of an interior transformation.

(3.1) *Let $T(A) = B$ be a 0-regular transformation of the continuum A into B. If K is a locally connected closed subset of B, then $T^{-1}(K)$ is locally connected.*

Since K is a locally connected closed subset of B, K can have only a finite number of components, each of which is a locally connected continuum. Let K' be any one of these and let C be any component of $T^{-1}(K')$, then $T(C) = K'$ is 0-regular.[7] Moreover, the number of components of $T^{-1}(K')$ must be finite, since each component of $T^{-1}(K')$ maps onto all of K' and for each point b of B the set $T^{-1}(b)$ is locally connected.[7] Thus in order to prove the theorem it suffices to show that C is locally connected. Since a 0-regular transformation is interior, this follows from (2.11) after establishing the following

(3.11) LEMMA. *Let $T(P) = Q$ be a 0-regular transformation of the continuum P into Q, then T has property β relative to (P, Q).*

Suppose the lemma false, then for some $\epsilon > 0$ there exist points a_n and a'_n

[7] See *Bulletin American Mathematical Society, abstract* 44-3-161.

[8] See Whyburn, "On sequences and limiting sets," *Fundamenta Mathematicae*, vol. 25 (1935), p. 412.

with $T(a_n) = T(a'_n) = b_n$ such that $\rho(a_n, a')_n < \epsilon/n$, while any connected subset K_n of A containing $a_n + a'_n$ has a diameter greater than ϵ. In case infinitely many of the b_n are distinct it may be assumed the sequence $\{b_n\}$ converges to some point b in B. But by supposition $\{T^{-1}(b_n)\}$ does not converge 0-regularly to $T^{-1}(b)$, which is contrary to the hypothesis that T is a 0-regular transformation. If almost all the $b_n = b$, then $T^{-1}(b)$ is not locally connected, which contradicts a known theorem.

It may be shown by simple examples that the 0-regular transformations are not the only interior transformations which have property β. However, the difficulties involved in obtaining conditions which insure property β for the general interior transformation is illustrated by the following

(3.2) EXAMPLE. *There exists a continuum A in the euclidean plane and a monotone interior transformation*[9] $T(A) = B$ *with the following properties:* (i) *for each point b of B the set $T^{-1}(b)$ is a simple arc,* (ii) *B is a simple arc, and* (iii) *A is not locally connected.*

In the euclidean plane define $f(x) = (1 + \sin 1/x)/2$ and let A be the closed portion of the plane bounded by the following sets: $y = f(x)$, $0 < x \leq 1$; $x = 0$, $0 \leq y \leq 1$; $y = 1$, $-1 \leq x \leq 0$; $x = -1$, $-1 \leq y \leq 1$; $y = -1$, $-1 \leq x \leq 1$; $x = 1$, $-1 \leq y \leq f(1)$. For each x_0, $0 < x_0 \leq 1$, define N_{x_0} as the sum of the following sets: $x = -x_0$, $-x_0 \leq y \leq 1$; $y = -x_0$, $-x_0 \leq x \leq x_0$; $x = x_0$, $-x_0 \leq y \leq f(x_0)$. Finally define N_0 to be the set $x = 0$, $0 \leq y \leq 1$. Clearly $\sum_{0 \leq x_0 \leq 1} N_{x_0} = A$. Let B be the unit interval, and for each point a of A, which must belong to some N_{x_0}, define $T(a) = x_0$. It will be observed that T fails to have property β near N.

(4) *Let A be a locally connected compact set and let $T(A) = B$ be a continuous transformation such that $T^{-1}(b)$ is locally connected for each point b of B. If K is a closed subset of B such that $F(B - K)$*[10] *is a finite number of points, then T has property β relative to $[T^{-1}(K), K]$.*

Suppose the theorem is not true. Then for some $\epsilon > 0$ and each n there exist points a'_n and a''_n of $T^{-1}(K)$ with $T(a'_n) = T(a''_n) = b_n$ such that $\rho(a'_n, a''_n) < \epsilon/n$ while no $a'_n + a''_n$ lies in a connected subset of $T^{-1}(K)$ of diameter $\leq \epsilon$. Furthermore, it may be assumed $\lim a'_n = \lim a''_n = a$ and $\lim b_n = b = T(a)$, a point of K, since A is compact, T is continuous and

[9] A transformation $T(A) = B$ is monotone interior provided it is interior and for each point b of B the set $T^{-1}(b)$ is connected.

[10] For any open set W, $F(W)$ denotes the set-theoretic boundary of W, i.e., the set $\bar{W} - W$.

K is a closed subset of the compact set B. For sufficiently large n there exist arcs $A_n = a'_n a''_n$ such that diameter of $A_n < d_n$ where $\lim d_n = 0$, since A is locally connected and $\lim \rho(a'_n, a''_n) = 0$. For n so large that $d_n < \epsilon$ the arcs A_n can not lie in $T^{-1}(K)$ by supposition. Hence there exists a sequence of point $\{p_{n_i}\}$ contained in $A - T^{-1}(K)$ with p_{n_i} a point of A_{n_i} and $\lim p_{n_i} = a$. Thus $\{T(p_{n_i})\}$ is contained in $B - K$ and $\lim T(p_{n_i}) = T(a) = b$. But b is a point of K, and consequently b is a point of $F(B - K)$.

Since $F(B - K)$ is a finite set and T is continuous there exists a number $\sigma < \epsilon/2$ such that $T[U(a, \sigma)] \cdot F(B - K) = b$. Moreover, for sufficiently large i the arc A_{n_i} lies in $U(a, \sigma)$. Thus $T(A_{n_i})$ is a continuum in $T[U(a, \sigma)]$ which contains a point in K and a point in $B - K$. Hence $T(A_{n_i})$ contains a point of $F(B - K)$, which must be the point b, i. e., $A_{n_i} \cdot F[A - T^{-1}(K)]$ is not empty and is contained in $T^{-1}(b)$. Let c'_i and c''_i be the first and last point respectively of A_{n_i} in $F[A - T^{-1}(K)]$. Then $\{c'_i\}$ and $\{c''_i\}$ are two sequences of points in $T^{-1}(b)$ such that $\lim \rho(c'_i, c''_i) = 0$. Now, by supposition and selection of σ, no pair of points c'_i and c''_i can lie together in a connected subset of $U(a, \sigma) \cdot T^{-1}(K)$. But $T^{-1}(b)$ is contained in $T^{-1}(K)$. Therefore $T^{-1}(b)$ is not locally connected, which is a contradiction.

(4.1) *If in the above theorem T is an interior transformation, then $T^{-1}(K)$ is locally connected.*

By the above theorem T has property β relative to $[T^{-1}(K), K]$. Moreover, $T[T^{-1}(K)] = K$ is interior by a known theorem.[11] Since T is continuous and A is a locally connected compact set, so also is B. Therefore K is locally connected, since $F(B - K)$ is finite. Thus the conditions of (2.11) are satisfied and, consequently, $T^{-1}(K)$ is locally connected.

(4.11) COROLLARY. *Let $T(A) = B$ be an interior transformation of the locally connected compact set A into a linear graph B such that $T^{-1}(b)$ is locally connected for every point b of B. If K is a subgraph of B, then $T^{-1}(K)$ is locally connected.*

(5) The continuous transformation $T(A) = B$ is *light*[2] provided that for each point b of B the set $T^{-1}(b)$ is totally disconnected. If $T^{-1}(b)$ is a finite (countable) set of points for each b, then T is a *finite* (*countable*) transformation.

(5.1) *Let $T(A) = B$ be a light transformation on the compact set A, let p be a point of A, and let Q be a closed subset of A contained in $A - p$. There*

[11] See Whyburn, "Interior transformations on compact sets," *loc. cit.*, (1.2), p. 370.

exists a number d such that every neighborhood W of $T(p)$ with $\delta(W) < d$ has the property that $\{T^{-1}[F(W)] - Q\}$ separates[12] *p and Q in A.*

If the component of A containing p does not intersect Q, then any neighborhood of $T(p)$ satisfies the theorem. Otherwise, there exists a neighborhood U of p such that $\bar{U} \cdot Q = 0$ and $F(U) \cdot T^{-1}[T(p)] = 0$ while $F(U) \neq 0$. Consequently, $T(p)$ and $T[F(U)]$ are non-empty disjoint compact subsets of B, since A is compact. Therefore $\rho[T(p), T\{F(U)\}] = d > 0$. Thus any neighborhood W of $T(p)$ with $\delta(W) < d$ has the property that $F(W)$ separates $T(p)$ and $T[F(U)]$ in B. Therefore $T^{-1}[F(W)]$ separates $T^{-1}[T(p)]$ and $T^{-1}[T\{F(U)\}]$ in A. In particular $T^{-1}[F(W)]$ separates p and $F(U)$ in A. Moreover, since $\bar{U} \cdot Q = 0$, the set $\{T^{-1}[F(W)] - Q\}$ must separate p and $F(U)$ in A. But $F(U)$ separates p and Q. Hence $\{T^{-1}[F(W)] - Q\}$ separates p and Q in A.

(5.2) *Let A be a continuum and let $T(A) = B$ be a finite (countable) transformation. If B is a regular (rational) curve, so also is A.*

In order to prove the theorem it suffices to show that if p and q are any two points of A, then p and q are separated in A by a finite (countable) set of points. Let d be the number given by the above theorem for p and q. Then, since B is a regular (rational) curve, there exists a neighborhood W of $T(p)$ such that $\delta(W) < d$ and $F(W)$ is a finite (countable) set of points. Hence $T^{-1}[F(W)]$ is a finite (countable) set, since T is finite (countable). But by the above theorem a subset of $T^{-1}[F(W)]$ separates p and q in A.

(6) A continuous transformation is *semi-monotone* if for every continuum K contained in B, the set $T^{-1}(K)$ has a finite number of components.

(6.1) *Let $T(A) = B$ be a light semi-monotone transformation of the compact set A. If B is a locally connected continuum, then A also is locally connected.*

Let p be any point of A and let K be any subcontinuum of $A - p$. In order to show that A is locally connected at p it suffices[13] to show that there exists a finite collection of continua $[g_i]$ such that $\Sigma g_i = A$ and for all i either $p \cdot g_i = 0$ or $K \cdot g_i = 0$. Let W be a neighborhood of $T(p)$ given by (5.1). Then $T(p)$ and $F(W)$ are disjoint closed subsets of the locally connected continuum B. Hence there exists[14] a finite collection of continua $\{h_j\}$ such

[12] A set K separates the two sets P and Q in A provided $K \cdot (P + Q) = 0$ and no component of $A - K$ intersects both P and Q.

[13] See Moore's "Foundations of Point Set Theory," Chapt. II, Theorem 51.

[14] See Moore, *loc. cit.*, Theorem 50.

that $\Sigma h_j = B$ and for each j either $T(p) \cdot h_j = 0$ or $F(W) \cdot h_j = 0$. But for each j, $T^{-1}(h_j)$ is a finite number of continua, since T is semi-monotone. Therefore $[T^{-1}(h_j)]$ is a finite collection of continua filling up A. Call this collection $[g_i]$. Now a subset of $T^{-1}[F(W)]$ separates p and K in A. Consequently, for each i either $p \cdot g_i = 0$ or $K \cdot g_i = 0$, since no h_j intersect both $T(p)$ and $F(W)$.

(6.2) *Let $T(A) = B$ be a finite interior transformation of a compact set A. If K is a locally connected closed subset of B, then $T^{-1}(K)$ is locally connected.*

Let Q be any component of K, then, since K is a locally connected compact set, Q is both open and closed in K. Hence $T^{-1}(Q)$ is both open and closed in $T^{-1}(K)$, since $T^{-1}(K)$ is compact. Thus in order to show that $T^{-1}(K)$ is locally connected it suffices to show that $T^{-1}(Q)$ is locally connected.

Now $T[T^{-1}(Q)] = Q$ is interior by a known theorem.[11] Thus, since Q is a continuum, the above theorem gives that $T^{-1}(Q)$ is locally connected because of the following

(6.21) LEMMA. *If $T(M) = N$ is a finite interior transformation of a compact set A, then T is semi-monotone.*

Let P be any subcontinuum of N, then for any component C of $T^{-1}(P)$, it follows from a known theorem that $T(C) = P$.[15] But T is finite. Hence the number of components, as C, must be finite. Therefore T is semi-monotone.

There are simple examples which show that the above theorem is not valid when the finite condition is removed. However, that finiteness is not enough to insure the invariance of local connectedness under the inverse of a continuous transformation is shown by the following

(6.3) EXAMPLE. *There exists a continuum A in the plane and a continuous transformation $T(A) = B$ with the following properties:* (i) *for each point b of B, $T^{-1}(b)$ consists of at most two points,* (ii) *B is locally connected, and* (iii) *A is not locally connected.*

Let A_1 be the closed unit square with two of its sides on the positive x- and y-axes respectively. Let q be the point $(0,0)$, $q_n = (-1/n, 0)$, $p = (0,1)$, and $p_n = (-1/n, 1)$. Let P be the segment from p_1 to p, Q the line segment from q_1 to q, and K_n the segment from q_n to p_n. Define $A_2 = P + Q + \Sigma K_n$, and $A = A_1 + A_2$. Let the transformation consist of folding A_2 onto A_1 about the y-axis.

UNIVERSITY OF VIRGINIA.

---

[15] See Whyburn, "Interior transformations on compact sets," *loc. cit.*, Theorem (1.5), p. 371.

SOME CHARACTERIZATIONS OF INTERIOR TRANSFORMATIONS.*

By A. D. WALLACE.

1. We suppose that S and S_1 are metric spaces and that S_1 is connected and non-degenerate. This is mainly to insure that every point of S_1 is a limit point of S_1, which is essential to some of the results. Let $T(S) = S_1$ be a single-valued continuous transformation. We say that T is *interior* if the image of every open set is open.[1] We do not require that the inverse of every point be totally disconnected. Except as noted we shall make use of the terminology of Kuratowski.[2] However if P is a point set we write $F(P) = \overline{P} \cdot \overline{CP}$, $P^0 = C\overline{CP}$, respectively the *boundary* and the *interior* of P, where CP denotes the complement of P. We call particular attention to the properties of the transformation T listed below.[3]

If Y is a subset of S_1 then $T^{-1}(CY) = CT^{-1}(Y)$.

If $\{Y_a\}$ is any collection of sets in S_1 then $T^{-1}(\sum_a Y_a) = \sum_a T^{-1}(Y_a)$ and

$$T^{-1}(\prod_a Y_a) = \prod_a T^{-1}(Y_a).$$

If $X \subset S$ and $Y \subset S_1$ then

$$T[X \cdot T^{-1}(Y)] = T(X) \cdot Y.$$

2. The first part of the following theorem is a generalization of a theorem stated by Eilenberg[4] for compact spaces.

(2.1) *Each of the following conditions is necessary and sufficient in order that $T(S) = S_1$ be interior:*

(i) *For each sequence of point-sets $\{Y_n\}$ converging to Y in S_1 we have* $\lim T^{-1}(Y_n) = T^{-1}(Y)$.

(ii) *For each set Y in S_1, $T^{-1}(\bar{Y}) = \overline{T^{-1}(Y)}$.*

(iii) *For each set X in S, $T(X^0) \subset T(X)^0$.*

---

* Received July 25, 1938; Revised October 21, 1938.

[1] S. Stoilow, *Leçons sur les principes topologiques de la théorie des fonctions analytiques*, Paris, 1938.

[2] C. Kuratowski, *Topologie I*, Warsaw, 1933.

[3] Kuratowski, *loc. cit.*, p. 12.

[4] S. Eilenberg, "Sur les transformations d'espaces métrique en circonférences," *Fundamenta Mathematicae*, vol. 24 (1935), p. 160.

Proof. We prove that interiority implies (i), (i) implies (ii), (ii) implies (iii), and (iii) implies interiority.

Assume that T is interior and let $Y = \lim Y_n$. We show first that $T^{-1}(Y)$ is a subset of $\liminf T^{-1}(Y_n)$. Let p be in the former set and let U be an open set containing p. If $U \cdot T^{-1}(Y_n)$ were empty for infinitely many n, then since $T[U \cdot T^{-1}(Y_n)] = T(U) \cdot Y_n$, the latter set would be vacuous for infinitely many n, which is impossible since $T(p)$ is in Y and $T(U)$ is an open set containing $T(p)$. Next, $\limsup T^{-1}(Y_n)$ is a subset of $T^{-1}(Y)$. If not, let p be a point of the former but not of the latter set. Then there is an open set V containing $T(p)$ and such that $V \cdot Y_n$ and hence $T^{-1}(V) \cdot T^{-1}(Y_n)$ is vacuous for all large n. But $T^{-1}(V)$ is open and contains p so that $T^{-1}(V) \cdot T^{-1}(Y_n)$ is not vacuous for infinitely many n. Hence we have shown that

$$\limsup T^{-1}(Y_n) \subset T^{-1}(Y) \subset \liminf T^{-1}(Y_n)$$

and the result follows.

To prove that (i) implies (ii) define Y_n to be Y. Then $\bar{Y} = \lim Y_n$ and $T^{-1}(Y_n)$ converges to $\overline{T^{-1}(Y)}$. Hence

$$T^{-1}(\bar{Y}) = T^{-1}(\lim Y_n) = \lim T^{-1}(Y_n) = \overline{T^{-1}(Y)}.$$

As to (iii) let X be a set in S. Then

$$\begin{aligned} T^{-1}[C(T(X)^0)] &= T^{-1}[\overline{CT(X)}] \\ &= \overline{T^{-1}CT(X)}, \text{ by (ii)} \\ &= \overline{CT^{-1}T(X)}, \text{ since } T^{-1} \text{ and } C \text{ permute.} \end{aligned}$$

Also $X \subset T^{-1}T(X)$ so that

$$CT^{-1}T(X) \subset \overline{CT^{-1}T(X)} \subset \overline{CX}.$$

But $\overline{CX} = C(X^0)$ and thus $T^{-1}[C(T(X)^0)] \subset CX^0$. Taking the complement of this inclusion and permuting T^{-1} and C we get

$$X^0 \subset T^{-1}[T(X)]^0$$

which implies (iii).

To show that (iii) implies interiority let $U = U^0$ be an open set, so that $T(U) = T(U^0) \subset T(U)^0$. But for any set P, $P^0 \subset P$. Thus $T(U) = T(U)^0$ which implies that $T(U)$ is open.

As a consequence of (ii) we have the following corollary.

(2.2) *If $T(S) = S_1$ is interior, for any $Y \subset S_1$*

(a) $$F[T^{-1}(Y)] = T^{-1}[F(Y)]$$

(b) $$T^{-1}(Y^0) = [T^{-1}(Y)]^0.$$

Proof. We have

$$T^{-1}[F(Y)] = T^{-1}[\bar{Y}\,\overline{CY}] = T^{-1}(\bar{Y}) \cdot T^{-1}(\overline{CY}) = \overline{T^{-1}(Y)} \cdot \overline{T^{-1}(CY)}$$
$$= \overline{T^{-1}(Y)} \cdot \overline{CT^{-1}(Y)} = F[T^{-1}(Y)]$$

where we use in turn the definition, the distributivity of T^{-1} over multiplication, (ii) of (2.1), the permutability of T^{-1} and C, and the definition. The proof of (b) is similar.

The results of the next corollary are typical rather than complete. There are quite a large number of descriptive point-set properties that are invariant under the inverse of an interior transformation. The proofs flow easily from (ii) of (2.1); the various definitions will be found in Kuratowski's book.[5]

(2.3) *The following properties are invariant under the inverse of an interior transformation,* i.e., *if Y in S_1 has the property so also has $T^{-1}(Y)$:*

(a) *The property of being a boundary set.*

(b) *The property of being non-dense.*

(c) *The property of being of the first category.*

Proof. (a) By the second part of (2.2) if $Y^0 = 0$ then $[T^{-1}(Y)]^0 = 0$.

(b) As in (a).

(c) If Y is of the first category then $Y = \sum_n Y_n$, where each Y_n is non-dense. Hence $T^{-1}(Y) = \sum_n T^{-1}(Y_n)$ is of the first category since each $T^{-1}(Y_n)$ is non-dense by (*b*).

(2.4) *The property of being of the second category is invariant under an interior transformation.*

Proof. If $X \subset S$ is of the second category and $T(X)$ is not, then $T^{-1}T(X)$ is of the first category by (2.3) part (c). But $X \subset T^{-1}T(X)$ so that we would have a set of the second category contained in a set of the first category. This is impossible.

3. An important class of interior transformations is the class of linear operations[6] mapping one Banach space *onto* (i.e. filling up) another Banach space. Conversely the interiority of such an operation is a powerful tool. In connection with this result and the problem of extending it to a wider class

[5] Kuratowski, *loc. cit.*, in particular Chapter I.

[6] S. Banach, *Théorie des opérations linéaires*, Warsaw, 1932, for the definitions. In particular Chapter III, theorem 3 (1).

of spaces we give an example to show that the theorem is *not* valid for non-separable, non-normed spaces. To this end let $M = [x]$, $x = (x_1, x_2)$, be the Cartesian vector plane re-metrized as follows:[7] If x and y are collinear with $\theta = (0, 0)$ let $\rho(x, y) = |x - y|$, where $|x|$ is the positive square root of $x_1^2 + x_2^2$. If x and y are not collinear with the origin let $\rho(x, y) = |x| + |y|$. Under this metric M is a connected, complete, linear space. For any $x \in M$ define $U(x) = x_1$, that is we project M onto the x_1-axis, M_1. Then M_1 is a Banach space and U is an additive, homogeneous and continuous operation. But U is *not* interior since there are open sets on the x_2-axis mapping onto θ, certainly not an open set. In particular the space M does not satisfy the first condition imposed on a space of type F,[8] i. e. we do not have in general $\rho(x, y) = \rho(x - y, \theta)$.

In connection with a theorem of the next section we note that the inverse of a point under a linear operation is not compact (if the operation is not 1 to 1). For if x and y are in the inverse of a point then so is the set L of all points $p = tx + (1 - t)y$, $-\infty < t < +\infty$, since we have

$$U(p) = tU(x) + (1 - t)U(y) = U(x) = U(y).$$

It is clear that the set L is not compact if it is in, say, a Banach space. We note moreover that the inverse of a point under a linear operation is connected (indeed convex) if the mapped space is a Banach space, and this was not the case in the example given above, where the inverse of every point except θ is totally disconnected. I am unable to say if this is pertinent to the problem, but it does indicate that the algebra and topology of M have parted company in an astonishing way.

4. We shall say that $T(S) = S_1$ is *strongly continuous* if for each sequence of sets $\{X_n\}$ converging to X in S we have $T(X) = \lim T(X_n)$. As a consequence of results of this section it may be seen that any continuous transformation on a compact space is strongly continuous. To the contrary, it may be seen that, using the last paragraph of the preceding section, no linear operation (unless it is 1 to 1) on a Banach space is strongly continuous, since by (4. 2) the inverse of each point must be compact.

For convenience we list together the following statements. In connection with III, see a condition placed by R. L. Moore[9] on his upper semi-continuous decompositions of spaces.

---

[7] Alexandroff and Hopf, *Topologie*, Berlin, 1935, p. 80, 5°.

[8] Banach, *loc. cit.*, p. 35, 1°.

[9] R. L. Moore, "Foundations of Point Set Theory," *American Mathematical Society Colloquium Publications*, vol. 13, p. 324. See also Theorem 6, p. 328.

I. *$T(S) = S_1$ is strongly continuous.*

II. *For each set $X \subset S$, $T(\bar{X}) = \overline{T(X)}$.*

III. *For each sequence $\{y_n\}$ of distinct points converging to $y \neq y_n$ in S_1, each infinite sequence $\{x_n\}$, $x_n \in T^{-1}(y_n)$, contains a subsequence converging to a point.*

(4.1) *If $T(S) = S_1$ is continuous then each of the above conditions implies the following. If in addition the inverse of every point is compact condition* III *implies condition* I.

Proof. That I implies II may be shown as in the proof of (2.1) that (i) implies (ii).

Assume now that II holds and that III does not. Then we can find a sequence of distinct points $\{y_n\}$ with $y_n \neq y = \lim y_n$ and a sequence $\{x_n\}$ with $x_n \in T^{-1}(y_n)$ and such that $X = \sum_n x_n$ contains no subsequence converging to a point, which implies that $X = \bar{X}$. Note also that $\bar{X}$ does not intersect the set $T^{-1}(y)$. Now we have

$$T(X) = T(\bar{X}) = \overline{T(X)} = \overline{\Sigma y_n} = y + \Sigma y_n.$$

Hence $y \in T(\bar{X})$ and $T^{-1}(y) \cdot \bar{X} \neq 0$. This is a contradiction.

We now assume that III holds and that the inverse of every point is compact. We let $X = \lim X_n$ and by continuity $T(X) \subset \liminf T(X_n)$. Let p be a point of $\limsup T(X_n)$. There is a sequence of points $\{p_{n_i}\}$, one from each set $T(X_{n_i})$, converging to p. Let $q_{n_i} \in X_{n_i} \cdot T^{-1}(p_{n_i})$. If infinitely many p_{n_i} are distinct we may select a subsequence of $\{q_{n_i}\}$ converging to a point q. If all but a finite number of the p_{n_i} are the same point then the corresponding q_{n_i} lie in the inverse of a point, a compact set by hypothesis, so that we may again select a subsequence converging to q. With no loss of generality we assume that $q_{n_i} \to q$. But $p_{n_i} = T(q_{n_i}) \to T(q)$ and hence $p = T(q)$. But $q_{n_i} \in X_{n_i}$ so that $q \in X = \limsup X_n$ and hence $p \in T(X)$. We have shown now that $\limsup T(X_n) \subset T(X)$. This completes the proof.

(4.2) *If $T(S) = S_1$ is interior and* III *holds then the inverse of every point is compact.*

Proof. Let $y \in S_1$ and (since S_1 is connected and non-degenerate) let $\{y_n\}$ be a sequence of distinct points converging to y. Let $\{q_n\}$ be any sequence of points chosen from $T^{-1}(y)$. Since by (2.1) $T^{-1}(y) = \lim T^{-1}(y_n)$, the 1-neighborhood U_1, of q_1, contains points from all but a finite number of the $T^{-1}(y_n)$ so that there exists a point $x_1 \in U_1 \cdot T^{-1}(y_{n_1})$. Similarly the

1/2-neighborhood, U_2, of q_2 contains points from all but a finite number of the sets $T^{-1}(y_n)$ and we may select a point $x_2 \in U_2 \cdot T^{-1}(y_{n_2})$. We secure in this manner a sequence $\{x_n\}$ with $x_n \in T^{-1}(y_{i_n})$ and $\rho(x_n, q_n) < 1/n$. As a consequence of III we may select a subsequence (we assume it to be the original sequence) $x_n \to q$, say. But clearly we have $q_n \to q \in T^{-1}(y)$ since this set is closed.

We shall say that T is *SCI* if it is both interior and strongly continuous. By way of recapitulation we state the following corollary:

(4.3) *Each of the conditions* II *and* III *is necessary and sufficient in order that the interior transformation* $T(S) = S_1$ *be SCI.*

Remarks. It is to be noted that condition II is equivalent to saying that the map of a closed set is closed. It is also interesting to note that the two conditions $T^{-1}(\bar{Y}) = \overline{T^{-1}(Y)}$ and $T(\bar{X}) = \overline{T(X)}$ imply a sort of compactness property of the space S, namely with respect to the sets $T^{-1}(y)$ as a collection and individually.

5. We assume in this section that T is *SCI*. We shall say that a set X in S is *normal* [10] if $T[F(X)]$ is a subset of $F[T(X)]$, that is to say, X is normal if all of its boundary points map into boundary points. The transformation T is *normal at the point* p in S if every neighborhood of p contains a normal neighborhood of p; it is *normal* if it is normal at every point; it is *light* [11] if the inverse of every point is totally disconnected.

(5.1) *For every set* X *in* S *we have* $F[T(X)] \subset T[F(X)]$.[11]

(5.2) *A set* X *is normal if and only if* $F[T(X)] = T[F(X)]$.

(5.3) *For any set* X *in* S *we have*

$$T(X)^0 - T(X^0) \subset T[F(X)] - F[T(X)] \subset T(X)^0.$$

Proof. Denote the members of the above inclusion by P, Q and R respectively. If p is in P it is not in $\overline{CT(X)}$ and not in $T(C\overline{CX})$. Hence $T^{-1}(p) \subset \overline{CX}$. Clearly $p \in T(X)$ so that $0 \neq T^{-1}(p) \cdot X \subset T^{-1}(p) \cdot \bar{X}$ and thus $T^{-1}(p) \cdot \bar{X} \cdot \overline{CX} \neq 0$ so that $p \in T[F(X)]$. Also p is not in $\overline{CT(X)}$ so that p is not in $\overline{T(X)} \cdot \overline{CT(X)} = F[T(X)]$. Hence $p \in P$ implies $p \in Q$. It is easily seen that $p \in Q$ implies $p \in R$.

(5.4) *If* X *is normal* $T(X^0) = T(X)^0$.[10]

---

[10] Stoilow, *loc. cit.*, chapter V.

[11] G. T. Whyburn, "Interior transformations on compact sets," *Duke Mathematical Journal*, vol. 3, no. 2, 1937, p. 370.

Proof. By (5.2) and (5.3) $T(X)^0 \subset T(X^0)$ and by (2.1, iii) we have the reverse inclusion.

(5.5) *For any normal open set X and any $y \in S_1$, if Y is a component of $T^{-1}(y)$ then $Y \subset X$ or $Y \cdot X = 0$.*

Proof. If $YX \neq 0 \neq Y - X$, since Y is connected $Y \cdot F(X) \neq 0$ and thus $y \in T[F(X)] = F[T(X)]$. But X is open so that $T(X)$ is open and we have $F[T(X)] = \overline{T(X)} - T(X)$. Since $Y \cdot X \neq 0$, y must be in $T(X)$. This is a contradiction.

(5.6) *If X is a normal open connected set then X is a component of $T^{-1}T(X)$.*

It may be seen by examples that if X is a normal connected open set not every component of $T^{-1}T(X)$ is normal, even when T is light.

(5.7) *If T is normal at p, then p is a component of $T^{-1}T(p)$.*

Proof. This follows from (5.5).

(5.8) *If T is normal it is light.*

This result is complementary to a result of Stoilow [10] and Whyburn [12] who show that if T is light and S is a 2-dimensional surface then T is normal.

THE UNIVERSITY OF VIRGINIA.

[12] G. T. Whyburn, "Interior transformations on surfaces," *American Journal of Mathematics*, vol. LX, no. 2, April, 1938, p. 477.

A METRIC CONNECTED WITH PROPERTY *S*.*

By J. L. KELLEY.

1. The following metric has been studied by Mazurkiewicz [1] and Urysohn [2] for locally connected continua and by G. T. Whyburn [3] for any connected, locally connected space. The distance $\rho^*(y, x)$ is defined to be the g. l. b. $[\delta(C)]$ where C is a connected set containing the points x and y and $\delta(C)$ is the diameter of C. This re-definition of metric leads to a space, homeomorphic with the original space, in which the spherical neighborhood of a point is connected. The question has been raised by G. T. Whyburn [4] as to the existence of a metric for a connected space having property S [5] such that the spherical neighborhood of a point will be connected and have property S. This paper contains the positive solution of this problem.

If we consider the transformation of metric $\rho(x, y)$ to $\rho^*(x, y)$ to be a transformation T of M into M^* the main result may be stated:

If M is a connected space having property S then M is homeomorphic with a space M^* having property S and in which the spherical neighborhood of a point is connected and has property S.

2. Definition of $\rho^*(x, y)$. We shall suppose M to be a connected, locally connected space. We now define a chain $R(\epsilon, x, y)$ as follows:

Definition. A collection of connected sets $A_1, A_2, \cdots A_n$ is a chain $R(\epsilon, x, y)$ provided that $x \in A_1$, $y \in A_n$, $A_iA_{i+1} \neq 0$ for $i = 1, \cdots n-1$ and *provided that the following inequalities are fulfilled*:

$$\delta(A_i) < \epsilon/4 \text{ for } i = 1, \cdots n \tag{1}$$

$$\epsilon - 2\delta(A_j) - 2\delta(A_i) - \sum_{k=j+1}^{i-1} \delta(A_k) > 0 \tag{2}$$

for all i, j, $1 \leq j < i \leq n$

---

* Received September 19, 1938.

[1] *Fundamenta Mathematicae*, vol. 1 (1920), pp. 167, 168.

[2] *Verhandelingen der Koninklyke Akademie von Wetenschappen te Amsterdam*, Erste Sectie, Deel XIII, No. 4 (1928), pp. 38-42.

[3] *American Journal of Mathematics*, vol. 54 (1932), pp. 367-376.

[4] *Ibid.*, pp. 536-538.

[5] A Metric space H is said to have property S provided that for each $\epsilon > 0$, H is the sum of a finite number of connected sets each of diameter $< \epsilon$. See Sierpinski, *Fundamenta Mathematicae*, vol. 1, pp. 44 and R. L. Moore, *ibid.*, vol. 3, p. 22.

We note that (2), by transposing the term $2\delta(A_i)$, is equivalent to

$$(2') \qquad \delta(A_i) < \tfrac{1}{2}\left[\epsilon - 2\delta(A_j) - \sum_{k=j+1}^{i-1} \delta(A_k)\right]$$

for each i, $1 \leq i \leq n$ and for every j, $1 \leq j < i$.

We now make the following definition:

Definition. The distance $\rho^*(x, y)$ between any two points x and y of M is defined to be the g. l. b of all numbers ϵ for which there exists a chain $R(\epsilon, x, y)$.

We must now show $\rho^*(x, y)$ has distance character. It may be shown without difficulty that the set of all points which lie with a given point x in some connected set of finite diameter is both open and closed in the locally connected, connected space M. If x and y lie together in a connected set of diameter D, this set constitutes a chain $R(\epsilon, x, y)$ with ϵ any number greater than $4D$. Hence, $\rho^*(x, y)$ is finite for any two points x and y. We must furthermore show:

(i) $\rho^*(x, y) = 0$ if and only if $x = y$.
(ii) $\rho^*(x, y) = \rho^*(y, x)$.
(iii) $\rho^*(x, y) + \rho^*(y, z) \geq \rho^*(x, z)$.

To prove statement (i) we first notice that $\rho^*(x, x) = 0$, for x is itself a chain $R(\epsilon, x, x)$ for any $\epsilon > 0$. Suppose now $\rho^*(x, y) = 0$. Then x, y lie together in a chain $R(\epsilon, x, y)$ for $\epsilon > 0$. If in (2) we set $j = 1$, $i = n$ we see that the diameter of the connected set $R(\epsilon, x, y)$ is less than ϵ. It follows that x and y are the same point. Statement (ii) follows from the symmetry in x and y of the definition of $\rho^*(x, y)$. To prove inequality (iii) it is sufficient to show that a chain $R(\epsilon_1, x, y)$ taken together with chain $R(\epsilon_2, y, z)$ forms a chain $R(\epsilon_1 + \epsilon_2, x, z)$. The requisite inequalities concerning the sets of $R(\epsilon_1 + \epsilon_2, x, z)$ are easily obtainable from the inequalities concerning the sets of $R(\epsilon_1 x, y)$ and $R(\epsilon_2, y, z)$.

3. *Properties of $\rho^*(x, y)$:*

THEOREM I. $\rho^*(x, y) \geq \rho(x, y)$

Proof. If in inequality (2) we set $j = 1$, $i = n$ we see that for a chain $R(\epsilon, x, y)$, ϵ is greater than the diameter of the set $R(\epsilon, x, y)$, which is greater than or equal to the distance $\rho(x, y)$. Hence every ϵ in the collection of numbers ϵ for which there exists a chain $R(\epsilon. x, y)$ is greater than $\rho(x, y)$ and $\rho^*(x, y) \geq \rho(x, y)$.

THEOREM II. *If C is a connected set $\delta(C) \leq \delta^*(C) \leq 4\delta(C)$, where $\delta^*(C)$ is the l. u. b. of $\rho^*(x, y)$ for all points x and y of C.*

Proof. The first part of this inequality follows from Theorem I. The second inequality may be shown if we note that for any two points x and y of C there is a chain $R(\epsilon, x, y)$, namely, C itself, with ϵ an arbitrarily small amount greater than $4\delta(C)$. Hence the distance $\rho^*(x, y) \leq 4\delta(C)$ and the inequality follows.

THEOREM III. *If M has property S so also has M^*; that is, if M has property S with respect to $\rho(x, y)$ M also has property S with respect to $\rho^*(x, y)$.*

This follows at once from Theorem II.

THEOREM IV. *The transformation $T(M) = M^*$ is a homeomorphism; that is, if a point p is a limit point of a set $P \subset M$ with respect to $\rho(x, y)$, it is also a limit point of P with respect to $\rho^*(x, y)$, and conversely.*

Proof. That limit points with respect to $\rho^*(x, y)$ are limit points with respect to $\rho(x, y)$ follows from the inequality $\rho(x, y) \leq \rho^*(x, y)$.

To prove the converse we note that a δ^* neighborhood of a point p (i. e. the set of all points x such that $\rho^*(p, x) < \delta$) is precisely the set of all points x for which there exists a chain $R(\delta, p, x)$. Now M is locally connected: hence there exists an arbitrarily small connected open set about any point p of M. A connected open set about p of diameter less than $\delta/4$ will be contained in a δ^* neighborhood of p, for such a set is itself a chain $R(\delta, p, x)$ where x is any point of the set. It follows that if p is a limit of a set $P \subset M$ with respect to $\rho(x, y)$ so also is it a limit point with respect to $\rho^*(x, y)$.

THEOREM V. *If M has property S, the δ^* neighborhood $V_{\delta^*}(N)$ of any subset $N \subset M$ has property S.*

Proof. The set $V_{\delta^*}(N)$ is precisely the set of all points x which can be joined to a point $y \in N$ by a chain $R(\delta, x, y)$. We shall show this set has property S with respect to $\rho(x, y)$.

Suppose l is an arbitrarily chosen number. We shall exhibit $V_{\delta^*}(N)$ as the sum of a finite number of connected sets each of diameter less than l. We assume $l/9 < \delta/4$.

Write $M = \sum_{i=1}^{s} B_i$, with each B_i connected and $\delta(B_i) < l/9$. Let $[B_{n_i}]$ be that subset of $[B_i]$ consisting of all members of $[B_i]$ which are contained

wholly in $V_{\delta^*}(N)$. Since $l/9 < \delta/4$ the sets B_{n_i} cover N. For each B_{n_i} let C_{n_i} be B_{n_i} together with every point x which lies with a point $y \in B_{n_i}$ in a connected subset of $V_{\delta^*}(N)$ of diameter less than $4l/9$. Each set C_{n_i} is connected and of diameter less than l. The proof will now be complete if we show that every point of $V_{\delta^*}(N)$ lies in some set C_{n_i}.

We next prove a Lemma:

LEMMA. *Suppose $A_1, A_2, \cdots A_p$ is a chain $R(\delta, x, y)$, with $x \in N$, and y belonging to a set B_i. Suppose further that*

$$l/9 < \tfrac{1}{2}\left[\delta - 2\delta(A_j) - \sum_{k=j+1}^{p} \delta(A_k)\right] \tag{3}$$

for $1 \leq j \leq p$. Then $B_i \subset V_{\delta^*}(N)$, i. e. B_i is a member of the collection $[B_{n_i}]$.

The truth of the Lemma follows from the fact that the set B_i may be taken as the link of index $p+1$ of a chain $R(\delta, x, z)$, $z \in B_i$, for statement (3) is the fullfilment of the requirement $(2)'$ for the $(p+1)$ link.

Suppose now, $y \in V_{\delta^*}(N)$ with y not a point of a set B_{n_i}. We show that y belongs to a set C_{n_i}. There exists a chain $R(\delta, x, y)$ of sets $A_1, \cdots A_n$, $x \in A_1 N$, $y \in A_n$. Choose the number p such that the inequality

$$\tfrac{1}{2}\left[\delta - 2\delta(A_j) - \sum_{k=j+1}^{i-1} \delta(A_k)\right] > l/9 \tag{4}$$

is satisfied for $1 \leq j < i \leq p+1$ but such that the inequality (4) is not satisfied for $i = p+2$ and for $j = L \leq p+1$. This number p exists for (4) is satisfied for $i = 1$, and if (4) were satisfied for all i, $1 \leq i \leq n+1$, from the Lemma it would follow that y was a point of some B_{n_i}. By use of the Lemma it is evident that every point of A_p is a point of some B_{n_i}. The proof will then be complete if we show that $\sum_{k=p+1}^{n} \delta(A_k) < 4l/9$, for the point y will then lie in some C_{n_i}. This inequality is shown in three steps:

(i) $L \neq p+1$. Suppose the contrary: Then $\frac{1}{2}[\delta - 2\delta(A_{p+1})] \leq l/9 < \delta/4$. But $\delta(A_{p+1}) < \delta/4$. Hence, $\delta - 2\delta(A_{p+1}) > \delta/2$, and we have a contradiction.

(ii) $\delta(A_{p+1}) < 2l/9$. We know that $\frac{1}{2}\left[\delta - 2\delta(A_L - \sum_{k=L+1}^{p+1} \delta(A_k)\right] \leq l/9$ and also since $L \neq p+1$ we may write $\delta - 2\delta(A_L) - 2\delta(A_{p+1}) - \sum_{k=L+1}^{p} \delta(A_k) > 0$. Multiplying the first inequality by two and subtracting the second term by term we have $\delta(A_{p+1}) < 2l/9$.

(iii) $\sum_{k=p+2}^{n} \delta(A_k) < 2l/9$. We know that

$$\tfrac{1}{2}\left[\delta - 2\delta(A_L) - \sum_{k=L+1}^{p+1} \delta(A_k)\right] \leqq l/9$$

and also $\delta - 2\delta(A_L) - 2\delta(A_n) - \sum_{k=L+1}^{n-1} \delta(A_k) > 0$.

Multiplying the first inequality by 2 and subtracting the second we have $2\delta(A_n) + \sum_{k=p+2}^{n-1} \delta(A_k) < 2l/9$. From this statement (iii) follows.

This completes the proof of the theorem.

COROLLARY 1. *If M has property S, the δ^* neighborhood of any point p has property S.*

COROLLARY 2. *If M is a locally connected continuum it may be so metricized that*

(1) The δ neighborhood of any set has property S and the closure of this neighborhood consists of a finite number of locally connected continua.

(2) The δ neighborhood of a point is connected, has property S and the closure of such a neighborhood is a locally connected continuum.

UNIVERSITY OF VIRGINIA.

ÜBER EINEN TAUBERSCHEN SATZ IM DREIKÖRPERPROBLEM.*

By J. KARAMATA.

Durch die Note des Herrn Boas, "A Tauberian theorem connected with the problem of three bodies," *American Journal of Mathematics*, vol. 61 (1939), pp. 161-164, angeregt, suchte ich nach einer notwendigen und hinreichenden KB. (Konvergenzbedingung), etwa von der Form der R. Schmidtschen KB.

$$\liminf_{x\to 0} \operatorname*{Min}_{x\leqq x'\leqq \lambda x} \{f'(x') - f'(x)\} \geqq -w(\lambda) \to 0, \qquad \lambda \to 1,$$

welche aber die Boassche KB.

$$xf''(x) = O\{\omega(|f(x)|)\}, \qquad x \to 0,$$

mit (für $y > 0$) positivem und wachsendem $\omega(y)$, enthalten sollte.

Wegen der etwas einfacheren Ausführung betrachte ich den Grenzübergang bei $x \to \infty$ statt bei $x \to 0$, und formuliere es als eine Umkehrung des $(C, 1)$-Verfahrens. Der Fall $x \to 0$ kann zwar nicht unmittelbar auf den Fall $x \to \infty$ zurückgeführt werden, doch kann eine vollständig analoge Überlegung auf den erstgenannten Fall angewendet werden.

Die besprochene Erweiterung lautet:

Aus
$$\frac{1}{x}\int_0^x s(t)\,dt \to s, \qquad x \to \infty, \tag{I}$$

folgt
$$s(x) \to s, \qquad x \to \infty, \tag{II}$$

wenn für eine in $(-\infty, +\infty)$ *stetige und eigentlich wachsende Funktion* $\phi(x)$ *die KB.*

$$\liminf_{x\to\infty} \operatorname*{Min}_{x\leqq x'\leqq \lambda x} \{\phi(s(x')) - \phi(s(x))\} \geqq -w(\lambda) \to 0, \qquad \lambda \to 1 \tag{III}$$

erfüllt ist.

Beweis. Aus (I) folgt erstens

$$\frac{1}{(\lambda-1)x}\int_x^{\lambda x} s(t)\,dt \to s, \qquad x \to \infty,$$

d. h. es gilt, von einem x ab,

$$s - \epsilon < \frac{1}{(\lambda-1)x}\int_x^{\lambda x} s(t)\,dt < s + \epsilon. \tag{IV}$$

* From a letter to Aurel Wintner, received April 15, 1939.

Zweitens folgt aus (III), von einem x ab,

$$\phi\{s(t)\} - \phi\{s(x)\} > -w(\lambda) - \epsilon' \text{ für } x \leqq t \leqq \lambda x,$$

d. h., wenn ψ die inverse Funktion von ϕ bedeutet,

$$s(t) > \psi\{\phi(s(x)) - w(\lambda) - \epsilon'\} \text{ für } x \leqq t \leqq \lambda x.$$

Also ist wegen der rechtsstehenden Ungleichung (IV)

$$s + \epsilon > \psi\{\phi(s(x)) - w(\lambda) - \epsilon'\},$$

d. h., von einem x ab,

$$\phi\{s(x)\} < \phi(s + \epsilon) + w(\lambda) + \epsilon',$$

woraus folgt, dass

$$\limsup_{x\to\infty} \phi\{s(x)\} \leqq \phi(s) \text{ ist.}$$

Entsprechend folgt aus der linksstehenden Ungleichung (IV) und der der KB. (III) äquivalenten Bedingung

$$\liminf_{x\to\infty} \operatorname*{Min}_{x\leqq x'\leqq\lambda x} \{\phi(s(\lambda x)) - \phi(s(x'))\} > -w(\lambda) \to 0, \qquad \lambda \to 1,$$

dass

$$\liminf_{x\to\infty} \phi\{s(x)\} \geqq \phi(s)$$

ist, woraus wegen der Stetigkeit und dem eigentlichen Wachstum von ϕ die Behauptung (II) folgt.

Aus der KB. (III) is nun leicht eine einfachere, hinreichende, der Boasschen Bedingung entsprechende KB. abzuleiten. Die KB. (III) is nämlich immer erfüllt, sofern nur

$$\phi'\{s(x)\}s'(x) > O(1/x), \qquad x \to \infty \tag{V}$$

gilt, mit $\phi'(y) > 0$ in $(-\infty, +\infty)$.

Man braucht nur (V) von x bis $x' \leqq \lambda x$ zu integrieren, um (III) mit $w(\lambda) = M \log \lambda$ zu erhalten.

Die KB. (V) stellt eben die der Boasschen KB. entsprechende einseitige KB. dar, mit $\phi'(y) = 1/\omega(|y|)$, woraus noch zu ersehen ist, dass die Voraussetzung der Monotonie über $\omega(y)$ überflüssig ist.

Man kann noch bemerken, dass in dem angeführten Satz die Voraussetzung (I) durch

$$\frac{1}{x}\int_0^x \Phi\{s(t)\}dt \to \Phi(s), \qquad x \to \infty,$$

ersetzt werden kann, wobei $\Phi(y)$ irgendeine in $(-\infty, +\infty)$ stetige und eigentlich wachsende Funktion bedeutet.

UNIVERSITY OF BELGRADE.

INFINITE NUMBER FIELDS WITH NOETHER IDEAL THEORIES.*

By SAUNDERS MAC LANE and O. F. G. SCHILLING.[1]

The abstract development of many different types of ideal theories in general rings, at the hands of W. Krull [2] and of other students of Emmy Noether suggests a natural converse problem: Can one characterize constructively the set of all rings in which a given type of ideal theory holds? In particular, in what rings does the complete Noether (or classical) prime ideal decomposition theory hold? For this second problem we obtain here a partial answer: if the elements of such a ring are integral in every valuation of the ring, then the ring is necessarily the set of algebraic integers from a classical (finite) algebraic number field or from a certain type of infinite algebraic number field. The other classical case of algebraic function fields is also considered, as well as the closely related problem of finding those fields for which every valuation [3] is discrete.

1. Fields with discrete valuations. We search for those fields K whose valuations exhibit the following simple properties:

(i) Divisor-discreteness: every valuation of K is discrete;

(ii) Divisor-finiteness: each element x of K has valuation $V(x) \neq 0$ in only a finite number of non-equivalent valuations V of K.

These conditions, suggested by the ordinary arithmetic theories, are certainly satisfied by the classical fields studied in arithmetic:

(I) Any finite algebraic number field;

(II) Any finite algebraic function field of one variable over an absolutely algebraic coefficient field Ω of prime characteristic.

Conditions (i) and (ii) are also satisfied by certain infinite algebraic fields, obtained by the adjunction of an infinite number of algebraic elements to the field of rationals or to a field $\Omega(t)$ of rational functions:

(I_∞) Any infinite algebraic number field K in which for each valuation V of K there is a finite algebraic subfield $K_V \subset K$ such that V is unramified and undecomposed in the extension from K_V to K. Consequently each valuation V of such a K has an infinite residue class field.

* Received November 14, 1938.

[1] Johnston Scholar at the Johns Hopkins University for the year 1938-39.

[2] W. Krull, "Idealtheorie," *Ergebnisse der Mathematik*, vol. 4, no. 3.

[3] "Valuation" is used throughout in the sense of an "exponential" non-archimedean valuation (of rank 1).

Since V is discrete in the finite field K_V, the unramified extension of V to K is discrete; hence condition (i). On the other hand, to prove the divisor-finiteness of such an infinite field we need only demonstrate that no $\mathfrak{p}$-adic valuation $V_\mathfrak{p}$ of the rational number field $R \subset K$ is "infinitely decomposed" in K. To say that a valuation V on a subfield of K is infinitely decomposed in K means that V has an infinite number of distinct extensions to K. Suppose that some $V_\mathfrak{p}$ is infinitely decomposed in K, and let $R = K_0 \subset K_1 \subset K_2 \subset K_3 \subset \cdots$ be a tower of finite algebraic number fields approximating K. Then some extension V_1 of $V_\mathfrak{p}$ to K_1 must be infinitely decomposed in K, some extension V_2 of V_1 to K_2 must again be infinitely decomposed in K, and so on. The sequence of infinitely decomposable valuations, $V_1, V_2, V_3, \cdots$, determines a valuation V_∞ of K, and this valuation V_∞ is certainly decomposed in the extension to K from any finite K_V, contrary to the conditions of (I ∞). Hence any field of type (I ∞) is divisor-finite.

The same divisor properties (i) and (ii) hold for the fields:

(II ∞) Any infinite algebraic function field K of one variable over an absolutely algebraic coefficient field Ω of prime characteristic, such that K is separable over $\Omega(t)$, for a suitably selected t in K, and such that for each valuation V of K there is a finite algebraic function field $K_V \subset K$, with V unramified and undecomposed in the extension from K_V to K.

In this case the residue class field of V in K must be an infinite extension of Ω. Hence such an infinite field K is possible only over a coefficient field Ω not algebraically closed. That such fields can be constructed will be established below (§ 4); for the present we consider the converse problem.

THEOREM 1. *If the field K is divisor-finite and divisor-discrete, as in conditions* (i) *and* (ii), *then K either has no non-trivial valuations and hence is an absolutely algebraic field of prime characteristic, or K is a finite or infinite number or function field of one of the types specified above in* (I), (I ∞), (II), *and* (II ∞).

Proof. Case 1. K has characteristic 0. If K is transcendental over its rational subfield R, then K has a non-discrete valuation. For if t in K is transcendental over R, any p-adic valuation of R can be extended to a valuation of the rational function field $R(t)$ in such a way that t has any prescribed value,[4] say $V(t) = \sqrt{2}$, $V(p) = 1$. An extension of this valuation to K is manifestly not discrete.

[4] See, for instance, S. Mac Lane, "A construction for absolute values in polynomial rings," *Transactions of the American Mathematical Society*, vol. 40 (1936), pp. 363-395; see Theorem 3. 1.

On the other hand, if K is an algebraic number field, and is not finite, as in type (I), each rational prime p can have $V(p) \neq 0$ in only a finite number of distinct valuations V of K. Therefore each p-adic valuation of R can decompose only a finite number of times and, by (i), ramify but a finite number of times in any generation of K. Therefore the ramification and decomposition is complete at some finite stage K_V, as asserted.

Case 2. K has characteristic $\chi \neq 0$. If K has transcendence degree 2 or more over its prime field, a non-discrete valuation of K is readily constructed. Hence K is an algebraic function field over Ω, the absolutely algebraic subfield of K. If K is not Ω, let t be an element transcendental over Ω and generate K by a tower

$$K_0 = \Omega(t) \subset K_1 \subset K_2 \subset K_3 \subset \cdots, \qquad K = \sum_{n=0}^{\infty} K_n, \tag{1}$$

where each extension K_{n+1}/K_n is algebraic of finite degree. If an infinite number of these extensions are inseparable, the field K would contain [5] all the roots $t^{p^{-1}}, t^{p^{-2}}, \cdots$ of the variable t. The valuation of $\Omega(t)$ obtained by setting $V(t) = 1$ then has an extension to K in which $V(t^{p^{-n}}) = 1/p^n$, contrary to the assumption that all valuations of K are discrete. Hence all the extensions must be separable beyond some point in the tower (1). The usual ramification and decomposition theory applies to these extensions, giving the results of (II ∞).

In this proof the condition of divisor-finiteness was used only in showing that each V can decompose only a finite number of times in successive extensions to K. The omission of this condition thus yields the following characterization of a class of fields:

THEOREM 2. *If every valuation of a field K is discrete, then K is either an absolutely algebraic field of prime characteristic or is a number or function field of one of the types* (I), (I ∞), (II), (II ∞), *where in the case of the infinite fields* (I ∞) *and* (II ∞), *it need now only be required that each valuation be unramified* (*but not necessarily undecomposed*) *in the extension from K_V to K.*

The analogous "relative" problem considers extensions K of a given base field k relative only to those valuations V of K which are trivial in k, i.e. that the value $V(a)$ is zero for every a of k. This description includes precisely the valuations (or prime divisors) used for an ordinary algebraic function field K with constant field k. A proof similar to that above gives the result:

---

[5] By a theorem of F. K. Schmidt, an inseparable extension of a field $\Omega(t)$ can always be made separable by using as variable some p^n-th root of t.

THEOREM 3. *If the field K is an extension of k such that every valuation V of K which is trivial on k is discrete and such that any x in K has $V(x) \neq 0$ in only a finite number of non-equivalent valuations of this type, then K is either an algebraic extension of k (in which case all valuations V trivial on k are trivial on the whole field) or K is a finite or infinite algebraic function field of one variable over k. If K is an infinite function field, then K is separable over some finite function field $K_0 \subset K$, and for each valuation V of K there is a finite function field $K_V \subset K$ such that V is unramified and undecomposed in the extension from K_V to K. Conversely, any separable and denumerably infinite function field K with such a K_V for each V has the specified relative divisor-finiteness and discreteness properties.*

In the converse, the assumption that K is denumerable (i. e., that K can be obtained by the successive adjunction to k of a denumerable set of elements) makes it possible to use a denumerable tower for K, as in the previous proof.

A similar relativization of Theorem 2 is possible. The infinite algebraic fields which are excluded in these theorems by the assumptions of finite ramification and decomposition still have valuations which are rational (a valuation V of K is rational if the associated group of values is a subgroup of the additive group of rational numbers). An inspection of the proof of Theorem 1 yields the following results:

THEOREM 4. *The fields K all of whose valuations are rational are*: (a) *absolutely algebraic fields of prime characteristic*; (b) *finite or infinite algebraic number fields*; (c) *finite or infinite algebraic function fields of one variable over an absolutely algebraic coefficient field of prime characteristic.*

2. Ideal decomposition in infinite rings. The rings of integers in the infinite fields considered above may have Noether ideal theories. A domain of integrity J is said to have a *Noether ideal theory* if the classical decomposition theorems for ideals in algebraic number fields hold in J. In other words, every integral ideal $\mathfrak{a}$ of J can be represented uniquely as a product of prime ideals, and furthermore, $\mathfrak{a} \subset \mathfrak{b}$ implies that $\mathfrak{a} = \mathfrak{b}\mathfrak{c}$ for some ideal $\mathfrak{c}$ of J.

THEOREM 5. *The ring of algebraic integers in an infinite number field K of type* (I ∞) *has a Noether ideal theory.*[6] *If K is an infinite algebraic function field of type* (II ∞) (*or of the relative denumerably infinite type* (II′∞) *described in Theorem* 3) *over a constant field Ω with $\chi(\Omega) \neq 0$, and if t in K is trancendental over Ω, then the ring J of all functions in K integral with respect to t has a Noether ideal theory.*

[6] The ideal theory of an infinite field K is not in general so simple; see Krull, "Idealtheorie in unendlichen algebraischen Zahlkörpern I and II, *Mathematische Zeitschrift*, vol. 29 (1928), p. 42 and vol. 31 (1930), p. 537.

The proof depends in both cases on the assumed finiteness of the decomposition and ramification of each prime ideal in the fields of type (I ∞) and (II ∞), so that these cases may be treated simultaneously thus:

LEMMA 1. *If the Noether ideal theory holds in each integral domain J_n of a tower $J_0 \subset J_1 \subset J_2 \subset \cdots$, if the quotient field K_n of each J_n is a finite separable algebraic extension of the previous quotient field K_{n-1}, and if for each prime ideal $\mathfrak{p}$ of J_0 there is an index n such that any prime ideal factor of $\mathfrak{p}$ in J_n is unramified and undecomposed in every extension J_m/J_n with $m > n$, then the composite $J = \Sigma J_n$ of the domains J_n has a Noether ideal theory, provided that each J_n is the integral closure of J_0 in K_n.*

Proof. For the Noether ideal theory to hold in a domain J, a set of necessary and sufficient conditions is:[7] (i) J is integrally closed in its quotient field; (ii) the ascending chain condition holds for ideals in J; (iii) every non-null prime ideal in J is divisorless. First, the integral closure of $J = \Sigma J_n$ in its quotient field $K = \Sigma K_n$ is immediate. Suppose, contrary to the chain condition, that there were an infinite ascending chain of ideals $\mathfrak{a}_1 < \mathfrak{a}_2 < \mathfrak{a}_3 < \cdots$ in J. The ideal $\mathfrak{a}_1 \cap J_0$ has a decomposition into prime ideals of J_0 each of which cannot ramify or decompose beyond a certain stage in the tower. Hence there is an index n such that each prime ideal factor $\mathfrak{p}$ of $\mathfrak{a}_1 \cap J_n = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_s^{e_s}$ is undecomposed and unramified beyond J_n. The chain of ideals $\mathfrak{a}_1 \cap J_n \leqq \mathfrak{a}_2 \cap J_n \leqq \mathfrak{a}_3 \cap J_n \leqq \cdots$ in J_n all have decompositions obtained from $\mathfrak{a}_1 \cap J_n = \mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_s^{e_s}$ by the omission of factors. There are only a finite number of such decompositions, and they remain the same in all subsequent rings J_m, so that the assumed infinite chain is impossible.

Suppose, finally, contrary to the third necessary condition, that a prime ideal $\mathfrak{p} \neq 0$ of J had a proper divisor, $\mathfrak{p} < \mathfrak{a} < J$. Then for some n, $\mathfrak{p} \cap J_n < \mathfrak{a} \cap J_n < J_n$, contrary to the properties of the prime ideal $\mathfrak{p} \cap J_n$ in J_n. This establishes the Lemma, and with it, Theorem 5.

3. Rings with Noether ideal theories. The set of all integral domains with Noether ideal theories is very extended, as it includes all valuation rings of discrete rank 1 valuations, quotient rings of domains with Noether ideal theories, and the like. We seek more specifically only for rings with an added "integral" property, where an element b of a ring is said to be *integral* in a valuation V if $V(b) \geqq 0$.

THEOREM 6. *If a domain of integrity J has a Noether ideal theory and if all elements of J are integral in every valuation of J, then J is an absolutely algebraic field of prime characteristic or is the ring of all algebraic integers*

[7] B. L. van der Waerden, *Moderne Algebra*, vol. II, p. 97.

in an algebraic number field K which is finite or infinite of type (I ∞). *Conversely, the ring of all integers in any such field* (I) *or* (I ∞) *has a Noether ideal theory and consists of elements integral in every valuation.*

The converse statement is known (for (I ∞), see Theorem 5). On the other hand, if J has a Noether ideal theory, each prime ideal $\mathfrak{p}$ generates a corresponding discrete $\mathfrak{p}$-adic valuation of J. If J consists only of integral elements, a known argument [8] shows that every valuation of J is equivalent to one of these $\mathfrak{p}$-adic valuations and hence that every valuation is discrete. The quotient field K of J is therefore divisor-discrete and divisor-finite in the sense of § 1, hence K is either absolutely algebraic of prime characteristic or is one of the fields (I), (II), (I ∞), (II ∞). In the absolutely algebraic case, K has no non-trivial valuations, hence no $\mathfrak{p}$-adic valuations, so the unit ideal J is the only ideal and $J = K$. The function fields (II) are impossible here, for in these fields only the constants are integral in every valuation. Since J contains the rational integers and is integrally closed, J must be the ring of all algebraic integers in its quotient field K, as asserted.

In the function-theoretic case one might similarly attempt to characterize the rings of integers in an algebraic function field (II) or (II ∞) as Noether rings J for which a finite number of valuations $V_1, \cdots, V_m$ can be so selected that every element of J is integral in every valuation not equivalent to one of the V_i. However, many other types of rings J with exactly these properties are possible. In the infinite function fields the assumption that the ring of integral functions of t has a Noether ideal theory does not insure that the poles of t will correspond to valuations which remain discrete in the infinite field. Other rings J with the same arithmetic properties may be found by starting from the ring J_0 of integers in an algebraic number field and taking as J a quotient ring, consisting of fractions whose denominators belong to a suitable multiplicatively closed set S of J which does not contain the zero element.

4. Construction of infinite examples. To show that infinite algebraic number fields with the properties used for type (I ∞) exist, take as a ground field any finite algebraic number field k, and let $\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3, \cdots$ be an enumeration of all prime ideals of k. For each prime ideal $\mathfrak{p}_i$ denote the corresponding $\mathfrak{p}_i$-adic valuation by V_i. A tower of finite algebraic number fields $k = K_0 \subset K_1 \subset K_2 \subset \cdots$ will next be constructed by an induction. Pick an integer $l_n > 1$ which is to be the degree of K_{n+1} over the preceding K_n. Let K_n be already constructed, and let $W_1, \cdots, W_r$ be the set of all extensions of the

[8] See, for instance, S. Mac Lane, "A construction for prime ideals," *Duke Mathematical Journal*, vol. 2 (1936), p. 492; footnote 3.

first $n+1$ valuations $V_1, \cdots, V_{n+1}$ of K_0 to K_n. Since the residue class field of each W_j in K_n is finite, there is for each j a monic polynomial $f_j(x)$ of degree t_n with coefficients in K_n and irreducible modulo W_j. By the Chinese remainder theorem, there is then a monic polynomial $f(x)$ of degree t_n with $f(x) \equiv f_j(x)$ modulo W_j for each j. This $f(x)$ is irreducible over K_n. The next field K_{n+1} of the tower is then to be obtained by adjoining to K_n a root of $f(x)$. In the extension K_{n+1}/K_n each W_j is then inert, for the defining equation $f(x)$ is irreducible modulo W_j. The infinite field $K = \Sigma K_n$ obtained by this tower is of type $(\mathrm{I}\,\infty)$, since any valuation V of K must be an extension of some valuation V_n of the ground field, and so is unramified and undecomposed beyond $K_V = K_n$.

For any infinite tower of fields $K = \Sigma K_n$ the formal product $m = \Pi t_n$ of all the degrees $t_n = [K_{n+1} : K_n]$ can be expressed as a Steinitz G-number $m = \Pi q_i^{e_i}$, in which the formal product may extend over an infinite number of rational primes q_i, each appearing to an exponent e_i which is $+\infty$ or a natural number. The G-degree $m = \Pi t_n$ of K/K_0 is independent of the tower used.[9] We have the result

THEOREM 7. *For any Steinitz G-number m there exists an infinite algebraic number field K of type $(\mathrm{I}\,\infty)$ and absolute degree m.*

A similar construction yields function fields of type $(\mathrm{II}\,\infty)$. Let the projected coefficient field Ω of characteristic p have as absolute degree a Steinitz G-number m (*i. e.,* let Ω be $GF(p^m)$). Our construction is possible if at each stage K_n the projected degree t_n is the degree of a potential extension of the residue class field $\mathbf{K}_n$ in each valuation W_j under consideration. Each such residue class field $\mathbf{K}_n$ is a finite extension of Ω, hence is of a type $GF(p^{m'})$, where $m' = ms$ for s a finite number. But a field of type $GF(p^{m'})$ has an extension of degree t if and only if[10] every prime factor of t appears in m' with a finite exponent. But m' has at least one such prime factor unless Ω is algebraically closed; hence

THEOREM 8. *There exists an algebraic function field K of type $(\mathrm{II}\infty)$ over every absolutely algebraic coefficient field Ω of prime characteristic and not algebraically closed. Furthermore such a K with an infinite Steinitz G-degree $m_0 = \Pi q_i^{\infty}$ exists if and only if the g. c. d. of m_0 and the degree m of Ω is a natural number.*

Here the infinite Steinitz G-degree m_0 of a function field K is obtained

---

[9] See invariant definition in O. F. G. Schilling, "Some remarks on class field theory . . . ," *American Journal of Mathematics*, vol. 59 (1937), p. 414.

[10] See § 16 in E. Steinitz, "Algebraische Theorie der Körper," ed. by R. Baer and H. Hasse, Leipzig and Berlin.

from the degree m' of K over a rational function subfield $\Omega(t)$ by dropping all primes with finite exponents. The resulting infinite degree is independent of the choice of t.

Our method of construction depends essentially on an enumeration of all valuations of the base field. The infinite function fields which are relatively divisor-discrete and finite over a coefficient field k, as in Theorem 3, cannot be treated by our method if k is not denumerable. A different device may then be used for fields k which can be obtained by purely transcendental extensions of a denumerable field Λ. For let Λ be a suitable denumerable field, x a variable over Λ, and K an infinite function field which is constructed by the previous method so as to be relatively divisor finite and discrete, hence of relative type (II′∞):

$$\Lambda(x) = K_0 \subset K_1 \subset K_2 \subset \cdots, \qquad K = \Sigma K_n. \tag{2}$$

Let T be a (non-enumerable) set of elements algebraically independent over K. Then the field $K' = K(T)$ is an infinite algebraic function field of one variable over the new constant field $\Lambda' = \Lambda(T)$. This new field is given by a tower

$$K_0' \subset K_1' \subset K_2' \subset \cdots, \qquad K' = \Sigma K_n', \tag{3}$$

where each $K_n' = K_n(T)$.

To show that K' is of relative type (II′∞), note first that the separability of K' over $K_0' = \Lambda'(x)$ is insured because the step by step construction of the original tower (2) can be so performed that each K_n is separable over K_0. It remains to prove that each valuation V' of K' trivial on Λ' is totally inert after a certain stage K_n'. It is known that this true for V' considered as a valuation V in the original tower (2).

Case 1. V is not trivial on K. As a function on $K' = K(T)$, V' applied to any polynomial $f(T) = \Sigma a_i t_1^{e_{1i}} t_2^{e_{2i}} \cdots t_m^{e_{mi}}$ with coefficients a_i in K and variables t_j in T gives

$$V'\Big(\sum_{i=0}^{n} a_i t_1^{e_{1i}} t_2^{e_{2i}} \cdots t_m^{e_{mi}}\Big) \geqq \text{Min}\,[V(a_0), \cdots, V(a_n)], \tag{4}$$

in consequence of the fact that V' is trivial on $\Lambda(T)$, so $V(t_j) = 0$. If the inequality sign in (4) holds for some $f(T) \neq 0$, we can divide $f(T)$ by one of its coefficients a_i of minimum value $V(a_i)$, obtaining a new polynomial $g(T) \neq 0$ with $V'[g(T)] > 0$ and with coefficients integral in K. In the residue class field $\mathbf{K}$ of K, $g(T)$ provides a non-trivial relation between the residue classes $\tau_1, \cdots, \tau_m$ of the variables $t_1, \cdots, t_m$. $\mathbf{K}$ is algebraic over Λ, hence $\tau_1, \cdots, \tau_m$ are algebraically dependent over Λ. However, V' is trivial on the coefficient field $\Lambda(T)$ so that the residue class field of K' must include

(a subfield isomorphic to) $\Lambda(T)$, in which the residue classes $\tau_1, \cdots, \tau_m$ are independent over Λ, a contradiction.

The equality in (4) therefore holds always; in consequence V' has the same value group as V and is, like V, discrete. Furthermore the valuation V' on any field $K_n' = K_n(T)$ of the new tower is obtained by the equation (4) from the valuation on the corresponding field K_n in such way that the residue class field $\mathbf{K}_n'$ of K_n' is a transcendental extension of the residue class field $\mathbf{K}_n$ by the algebraically independent residue classes τ. If n is chosen so large that V is totally inert in each subsequent extension of the original tower (2), V' must therefore behave similarly in (3), so that it is subject to no further ramification or decomposition.

Case 2. V is trivial on K. First a reduction to the situation when only one t is adjoined! If V' is not altogether trivial, there must be an irreducible monic polynomial $\Phi(x)$ in $\Lambda'[x]$ with $V'(\Phi(x)) \neq 0$. Since $V'(x) = 0$, $V'(\Phi(x)) > 0$. At least one coefficient b of Φ is not in Λ, for V was assumed trivial on $K \supset \Lambda(x)$. Let t be one of the variables of T on which b is essentially dependent;[11] i. e., b is not algebraic over the field $\Lambda^* = \Lambda(T^*)$ obtained by adjoining the set $T^* = T - \{t\}$. We now investigate this chain of coefficient fields

$$\Lambda \subset \Lambda^* = \Lambda(T^*) \subset \Lambda' = \Lambda^*(t) \tag{5}$$

On $\Lambda^*(x)$, and hence on its algebraic extension $K^* = K(T^*)$, V' must be trivial. Otherwise there would be an irreducible monic polynomial $\Phi^*(x)$ in $\Lambda^*[x]$ with $V'(\Phi^*(x)) > 0$. The coefficient b of $\Phi(x)$ is not in Λ^*, so cannot figure in $\Phi^*(x)$. The g. c. d. of $\Phi(x)$ and $\Phi^*(x)$ is then 1, for both are irreducible over $\Lambda' = \Lambda^*(t)$. This g. c. d. $1 = a(x)\Phi(x) + b(x)\Phi^*(x)$ is impossible because $V'(\Phi) > 0$, $V'(\Phi^*) > 0$. Hence V is trivial on K^*.

K^* can be approximated by a tower of fields $K^*_n = K_n(T^*)$ corresponding to the original tower (2):

$$\Lambda^*(x) = K^*_0 \subset K^*_1 \subset K^*_2 \subset \cdots, \qquad K^* = \Sigma K^*_n. \tag{6}$$

The whole field $K' = K(T) = K(T^*, t) = K^*(t)$ is a simple transcendental extension of K^*, and the valuation V' trivial on K^* is accordingly determined[12] on K' by an irreducible polynomial $\psi(t)$ with $V'(\psi) > 0$. Specifically, $V'(g\psi^m) = m\lambda$ for any polynomial $g(t)$ prime to $\psi(t)$ and for $\lambda = V'(\psi)$ a constant. All coefficients of $\psi(t)$ lie in some field K^*_m of the tower (6), so the same polynomial $\psi(t)$ determines V' in each of the fields $K^*_n(t) = K'_n$ for which $n \geqq m$. Each residue class field $\mathbf{K}'_n$ of K'_n is then $\mathbf{K}'_n =$

[11] E. Steinitz, *loc. cit.*, § 22, number 7.

[12] This holds if $V'(t) \geqq 0$. If $V'(t) < 0$, we replace t by $1/t$.

$K^*_n(\Theta)$, for Θ a root of $\psi(t) = 0$. The degrees $[\mathbf{K}'_n : \mathbf{K}^*_n] = [\mathbf{K}^*_n(\Theta) : \mathbf{K}^*_n]$ are then the same for all $n \geqq m$; hence

$$[\mathbf{K}'_{n+1} : \mathbf{K}'_n] = [\mathbf{K}'_{n+1} : \mathbf{K}^*_{n+1}] \cdot [\mathbf{K}^*_{n+1} : \mathbf{K}^*_n] \,/\, [\mathbf{K}'_n : \mathbf{K}^*_n] = [\mathbf{K}^*_{n+1} : \mathbf{K}^*_n].$$

On the other hand, the adjunction of the transcendent t to K^*_n leaves the degree of the quantity defining K^*_{n+1} over K^*_n unaltered, so $[K'_{n+1} : K'_n] = [K^*_{n+1} : K^*_n]$. Combined with the previous result, this gives

$$[\mathbf{K}'_{n+1} : \mathbf{K}'_n] = [K'_{n+1} : K'_n] \qquad (n \geqq m).$$

In other words, the whole degree of the finite algebraic extension K'_{n+1}/K'_n appears as the degree of the corresponding residue class field extension $\mathbf{K}'_{n+1}/\mathbf{K}'_n$, which means that V' can be neither ramified nor decomposed in this extension, as was asserted. We have proven

THEOREM 9. *If Λ' is a field obtained by any purely transcendental extension of a finite field or of a finite algebraic number field, then there is over the coefficient field Λ' an infinite function field K', with arbitrary infinite Steinitz G-degree, of the relative type* (II'_∞) *described in Theorem* 3; *that is, K' is separable over a suitable purely transcendental subfield $\Lambda'(t)$ and each valuation V' of K' is unramified and undecomposed over a suitable finite algebraic extension of $\Lambda'(t)$.*

5. A connection between ideal-theories and valuations. The essential feature of the arithmetic of the function fields is the presence of a Noether ideal theory in the ring J of integral functions relative to every possible independent variable t. This leads to the following ideal-theoretic characterization of a class of fields:

THEOREM 10. *A field K is divisor-discrete and divisor-finite if and only if every integrally closed domain of integrity $J \subset K$ has a Noether ideal theory.*

Proof. Suppose first that each J has such an ideal theory. Then K is divisor-discrete. For otherwise some valuation V of K is non-discrete, so the value group of V contains a descending sequence of positive real numbers $\alpha_1 > \alpha_2 > \alpha_3 > \cdots > 0$. The valuation ring B_V of V is integrally closed, hence has a Noether ideal theory. However, each value α_n determines in B_V an ideal $\mathfrak{a}_n$ consisting of all elements a_n of K with $Va_n \geqq \alpha_n$. The resulting infinite ascending chain of ideals $\mathfrak{a}_1 < \mathfrak{a}_2 < \mathfrak{a}_3 < \cdots$ contradicts the known Noether ideal theory of B_V.

Secondly, if K were not divisor-finite, there would be in K an element x with $Vx > 0$ for an infinite set of non-equivalent valuations V. The intersection $J = \Delta B_V$ of the valuation rings corresponding to all of these valuations

V is then an integrally closed ring. Therefore the element x has a finite decomposition into prime ideals in J. By the usual proof,[8] every valuation W of J in which all elements of J are integral is a $\mathfrak{p}$-adic valuation determined by some prime ideal $\mathfrak{p}$ of J. Thus x has $W(x) > 0$ only in the finite set of non-equivalent $\mathfrak{p}$-adic valuations W corresponding to prime ideal factors of x, counter to the assumption that x has $Vx > 0$ for an infinite set of V's.

Conversely, suppose that K is divisor-discrete and divisor-finite, so that K is a field of one of the types enumerated in Theorem 1. If K is absolutely algebraic of prime characteristic, then any integrally closed J is K itself, so has a trivial Noether ideal theory. Secondly, if K is a finite or infinite number field, J must necessarily include the ring J_0 of all algebraic integers of K, and this ring J_0 is known to have a Noether ideal theory (Theorem 5). A similar situation arises if K is a function field of type (II) or (II$_\infty$) over an absolutely algebraic field Ω. In this case either $J = \Omega$, and the Noether ideal theory in J is trivial, or $J > \Omega$, so that there is an element t in J not in Ω. This t is then transcendental over Ω, and the integrally closed ring J must contain all functions in K integral with respect to t. The ring J_0 of these integral functions again has a Noether ideal theory, by Theorem 5. In both the number-theoretic and the function-theoretic case the required ideal theory in J may be obtained from that in the smaller domain J_0 by the following Lemma, which was proven for the case of algebraic number fields K by Skolem.[13]

LEMMA 2. *If a domain of integrity J_0 has a Noether ideal theory and if J is another domain of integrity containing J_0 and contained in the quotient field of J_0* (i.e., $J_0 \subseteq J \subseteq K$), *then J has a Noether ideal theory.*

Proof. If J contains an element θ of K with $(\theta) = \mathfrak{a}/\mathfrak{b}$, for $\mathfrak{a}$ and $\mathfrak{b}$ relatively prime ideals of J_0, then J contains any element γ of the form $(\gamma) = \mathfrak{c}/\mathfrak{b}^n$, for $\mathfrak{c}$ an ideal of J. This fact may be proven, following Skolem,[13] as follows. Pick an ideal $\mathfrak{d}$ such that $\mathfrak{b}^n\mathfrak{d}$ is a principal ideal [7] (ω) of J_0. Then

$$\mathfrak{c} = (\gamma)\mathfrak{b}^n, \qquad \mathfrak{c}\mathfrak{d} = (\gamma)\mathfrak{b}^n\mathfrak{d} = (\gamma\omega),$$
$$\mathfrak{a}^n = (\theta^n)\mathfrak{b}^n \qquad \mathfrak{a}^n\mathfrak{d} = (\theta^n)\mathfrak{b}^n\mathfrak{d} = (\theta^n\omega).$$

Both $\theta^n\omega$ and ω are integers (elements of J_0), and their g. c. d. is

$$(\theta^n\omega, \omega) = (\mathfrak{a}^n\mathfrak{d}, \mathfrak{b}^n\mathfrak{d}) = (\mathfrak{a}^n, \mathfrak{b}^n)\mathfrak{d} = \mathfrak{d}.$$

But the integer $\gamma\omega$ is in $\mathfrak{d}$, so there must exist integers ξ and η with

$$\gamma\omega = \xi\theta^n\omega + \eta\omega, \qquad \gamma = \xi\theta^n + \eta.$$

---

[13] Th. Skolem, "Integritätsbereiche in algebraischen Zahlkörpern," *Skrifter utgit av videnskapsselskapet*, Oslo, 1921, Nr. 21, especially Theorems 11 and 23.

By hypothesis, $\theta \in J$; by construction $\xi, \eta \in J_0 \subset J$. Hence $\gamma \in J$, as asserted above.

For any prime ideal $\mathfrak{p}$ of K let $B_{\mathfrak{p}}$ denote the ring of all elements θ of K which can be represented as ideal quotients $(\theta) = \mathfrak{a}/\mathfrak{b}$ with an ideal denominator $\mathfrak{b}$ prime to $\mathfrak{p}$. This $B_{\mathfrak{p}}$ is simply the valuation ring for the $\mathfrak{p}$-adic valuation $V_{\mathfrak{p}}$ of K. Let S denote the set of those prime ideals $\mathfrak{p}$ of K for which $J \subset B_{\mathfrak{p}}$. Then we assert that J is the intersection

$$J = \underset{\mathfrak{p} \in S}{\Delta} B_{\mathfrak{p}}.$$

We need only prove [14] that any element a in the intersection must lie in J. Such an element a is an ideal quotient

$$(a) = \mathfrak{a}(\mathfrak{p}_1^{e_1}\mathfrak{p}_2^{e_2} \cdots \mathfrak{p}_t^{e_t})^{-1}, \qquad \mathfrak{p}_1, \cdots, \mathfrak{p}_t \text{ not in } S.$$

For each denominator prime ideal $\mathfrak{p}_i$, J is not contained in $B_{\mathfrak{p}_i}$, so that there must be in J an element with ideal denominator divisible by $\mathfrak{p}_i$, and hence an element $(\gamma_i) = \mathfrak{c}_i/\mathfrak{p}_i$ whose ideal numerator $\mathfrak{c}_i$ has $(\mathfrak{c}_i, \mathfrak{p}_i) = 1$. By adding to γ_i, if necessary, a suitable element of J_0 we can assume that $(\mathfrak{c}_i, \mathfrak{p}_j) = 1$ for every j. The element

$$\gamma_1^{e_1} \cdots \gamma_t^{e_t} = (\mathfrak{c}_1^{e_1} \cdots \mathfrak{c}_t^{e_t})\,(\mathfrak{p}_1^{e_1} \cdots \mathfrak{p}_t^{e_t})^{-1}$$

of J then has numerator and denominator relatively prime. Its denominator is that of (a), so by the result of Skolem given above, a itself is in J.

The ring $J = \Delta B_{\mathfrak{p}}$ is now known to consist of those elements of K whose ideal denominators contain no prime factor from the set S. Such a ring J is a *regular quotient ring* of J_0 in the sense of Krull.[15] Theorem 3 of Krull's paper establishes a one-to-one correspondence between all the ideals $\mathfrak{b}_J$ of J and the ideals $\mathfrak{b}$ of J_0 whose prime ideal factors are all in the set S. This one-to-one correspondence $\mathfrak{b}_J \leftrightarrow \mathfrak{b}$ is given either by $\mathfrak{b}_J = \mathfrak{b} \cdot J$ or by $\mathfrak{b} = \mathfrak{b}_J \cap J_0$. ($\mathfrak{b}_J$ is the so-called extended ideal of $\mathfrak{b}$ to J.) Under this correspondence a prime ideal of J always corresponds to a prime ideal of J_0, and the inclusion relations between ideals of J are reproduced between the corresponding ideals of J_0. One then can show readily that the three sufficient conditions [7] for a Noether ideal theory in J are satisfied. We have completed the proof of Lemma 2 and, with it, of Theorem 10.

HARVARD UNIVERSITY,
THE JOHNS HOPKINS UNIVERSITY.

---

[14] Theorem 10 in Skolem, *loc. cit.*, is substantially equivalent to this assertion.

[15] W. Krull, "Zur Theorie der allgemeinen Zahlringe," *Mathematische Annalen*, vol. 99 (1928), pp. 51-70. Compare also W. Krull, "Idealtheorie," p. 17 and H. Grell, "Beziehungen zwischen den Idealen verschiedener Ringe," *Mathematische Annalen*, vol. 97 (1927), pp. 490-523.

RING HOMOMORPHISMS WHICH ARE ALSO LATTICE HOMOMORPHISMS.*

By MORGAN WARD.

1. Given two homomorphic rings $\mathfrak{D}$ and $\mathfrak{D}'$: what lattice properties of the rings are preserved under the homomorphism; more specifically, if $\mathfrak{D}$ is a lattice, will $\mathfrak{D}'$ be a homomorphic lattice?[1] It is easily seen that if $\mathfrak{S}$ and $\mathfrak{S}'$ are the lattices of ideals of $\mathfrak{D}$ and $\mathfrak{D}'$, any ring homomorphism of $\mathfrak{D}$ to $\mathfrak{D}'$ induces a lattice homomorphism of $\mathfrak{S}$ to $\mathfrak{S}'$. Unfortunately, when $\mathfrak{D}$ is a lattice with respect to the usual division relation, it need not be a sublattice of $\mathfrak{S}$. The homomorphism $\mathfrak{S}$ to $\mathfrak{S}'$ consequently gives little information about the lattice properties of $\mathfrak{D}'$.

If we assume however that the ascending chain condition holds in the lattice $\mathfrak{D}$, it is not difficult to show that $\mathfrak{D}$ is a sublattice of $\mathfrak{S}$ if and only if every ideal of $\mathfrak{D}$ is principal and $\mathfrak{D}$ and $\mathfrak{S}$ are lattice isomorphic. The object of this paper is to show in this case that all homomorphic rings $\mathfrak{D}'$ are also lattices of a very simple structure.

For the case when $\mathfrak{D}$ is a domain of integrity, my results may be more easily obtained from the fact that the fundamental theorem of arithmetic holds in every "principal ideal ring." (van der Waerden 1). The interest of the present investigation is that $\mathfrak{D}$ is merely required to be a commutative ring with a unit element. The methods used are based upon a theory of residuated lattices which have been developed by Mr. R. P. Dilworth and myself in a series of recent papers.[2]

2. Let $\mathfrak{D}$ be a commutative ring with a unit element, all of whose ideals are principal.

DEFINITION 2.1. DIVISION IN $\mathfrak{D}$. *If a and b are any two elements of $\mathfrak{D}$, a is said to divide b if and only if the principal ideal (a) contains the principal ideal (b). If (a) equals (b), a and b are said to be equivalent.*

We write $a \supset b$, $a \sim b$. $\mathfrak{D}$ evidently forms a semi-ordered set with respect

* Received December 16, 1938.

[1] This problem was propounded to me by Professor E. T. Bell for the special case when $\mathfrak{D}$ is the ring of rational integers.

[2] Ward, 1, 2; Ward-Dilworth, 1, 2.

to the relation $x \supseteq y$, and $x \sim y$ is an equivalence relation. Furthermore $a \supset b$ if and only if there exists an element c such that $ac = b$.

THEOREM 2.1. *The ascending chain condition holds for the elements of* $\mathfrak{O}$.

That is, if we have a chain of elements $a_1, a_2, a_3, \cdots$ such that $a_1 \subset a_2 \subset a_3 \subset \cdots$, then from a certain point on, all elements are equivalent to one another. It obviously suffices to show that in the ascending chain of ideals $(a_1) \subset (a_2) \subset (a_3) \subset \cdots$ all ideals are equal from a certain point on. The proof may be carried out exactly as in van der Waerden 1, § 17.

THEOREM 2.2. *The ideals of* $\mathfrak{O}$ *form a distributive residuated lattice.*

Proof. The ideals of any ring form a modular lattive $\mathfrak{S}$ over which a multiplication may be defined with the properties given in Ward 1. Since the ascending chain condition holds for ideals by the previous theorem, a residual may also be introduced, so that the lattice is residuated by definition. (Ward-Dilworth 1.) Since all ideals $\mathfrak{a}$, $\mathfrak{b}$ of $\mathfrak{O}$ are principal, $\mathfrak{a}$ contains $\mathfrak{b}$ if and only if there exists an ideal $\mathfrak{i}$ such that $\mathfrak{a}\mathfrak{i} = \mathfrak{b}$. Hence the lattice is distributive. (Ward-Dilworth 2, theorem 16.2).

3. We next consider the lattice formed by the elements of $\mathfrak{O}$.

THEOREM 3.1. *If a and b are any two elements of $\mathfrak{O}$, there exists an element d with the properties*

(i) $d \supset a$, $d \supset b$.
(ii) $x \supset a$, $x \supset b$ *imply* $x \supset d$.
(iii) $d = ua + vb$ *for some* u, v *of* $\mathfrak{O}$,
(iv) d *is determined up to equivalence.*

Proof. Consider the ideal $((a), (b)) = (a, b)$. Since all ideals are principal, $(a, b) = (d)$. Hence (i) follows and (iii) follows. Then (ii) follows from (iii), and (iv) from (ii).

We write $a \sim (a, b)$, and call d a union of a and b.

THEOREM 3.2. *If a and b are any two elements of $\mathfrak{O}$, there exists an element m such that*

(i) $a \supset m$, $b \supset m$.
(ii) $a \supset y$, $b \supset y$ *imply* $m \supset y$.
(iii) m *is determined up to equivalence.*

Proof. Consider the ideal $[(a), (b)]$. Since it is principal, $[(a), (b)] =$

(m). The element m is easily seen to have the required properties. We write $m \sim [a, b]$, and call m a cross-cut of a and b.

It follows from theorem 3.1 and 3.2 that $\mathfrak{D}$ forms a lattice with respect to the division relation $x \supset y$.

THEOREM 3.3. *$\mathfrak{D}$ forms a residuated lattice with respect to division relation $x \supset y$ and the multiplication operation xy of $\mathfrak{D}$.*

Proof. We need only show that the multiplication has the properties given in Ward 1, p. 629, for then by theorem 3.2, a residual may be introduced (Ward 1) so that $\mathfrak{D}$ will be residuated by definition. All of these properties are evident save the rule $a(b, c) \sim (ab, ac)$. Now $(b, c) = bu + cw$ by theorem 3.1 (iii) for some u, w of $\mathfrak{D}$. Hence $a(b, c) = abu + acw$. Therefore $(ab, ac) \supset a(b, c)$. But since $(b, c) \supset b$ and c, $a(b, c) \supset ab$, ac. Hence by theorem 3.1 (ii), $a(b, c) \supset (ab, ac)$. Therefore $(ab, ac) \sim a(b, c)$.

THEOREM 3.4. *$\mathfrak{D}$ is a distributive lattice.*

Proof. The correspondence $a \rightarrow (a)$ is a lattice isomorphism, since "equal elements" (that is equivalent elements in $\mathfrak{D}$) correspond to equal elements in $\mathfrak{S}$ and vice versa. Since $\mathfrak{S}$ is distributive by theorem 2.3, $\mathfrak{D}$ is distributive.

THEOREM 3.5. *Every element of $\mathfrak{D}$ may be uniquely represented up to equivalence as a cross-cut of primary elements none of which divides any other.*

Proof. With the terminology of Ward-Dilworth 2, $\mathfrak{D}$ is a distributive residuated lattice in which every element is principal. The result thus follows from theorem 14.2 of Ward-Dilworth 2, theorem 8.4 of Ward 2 and the remarks in section 12 of Ward-Dilworth 2.

4. Consider any homomorphism of the ring $\mathfrak{D}$. The homomorphism is completely specified by an ideal $\mathfrak{m}$, and its residue classes. Since $\mathfrak{m}$ is a principal ideal (m), $a \equiv b \pmod{\mathfrak{m}}$ if and only if $a = b + qm$ for some element q of $\mathfrak{D}$. Denote the residue classes of congruent elements modulo $\mathfrak{m}$ by $A, B, \cdots$. We wish to make these classes into a lattice. We begin by extending definition 2.1.

DEFINITION 4.1. DIVISION MODULO $\mathfrak{m}$. *An element a of $\mathfrak{D}$ is said to divide another element b of $\mathfrak{D}$ modulo $\mathfrak{m}$ if and only if there exists an element c such that $ac \equiv b \pmod{\mathfrak{m}}$.*

We write $a \supset b \pmod{\mathfrak{m}}$.

DEFINITION 4.2. EQUIVALENCE MODULO $\mathfrak{m}$. *Two elements a and b of $\mathfrak{D}$ are said to be equivalent modulo $\mathfrak{m}$ if and only if each divides the other modulo $\mathfrak{m}$.*

We write $a \sim b \pmod{\mathfrak{m}}$.

$\mathfrak{D}$ forms a semi-ordered set with respect to the relation of division modulo $\mathfrak{m}$, and equivalence modulo $\mathfrak{m}$ is an equivalence relation in the technical sense. It is evident furthermore that we have

THEOREM 4.1. *$a \supset b$ modulo $\mathfrak{m}$ if and only if there exist elements r and s such that $b = ar + ms$.*

We may note also that $a \sim b$ or $a \equiv b \bmod \mathfrak{m}$ imply $a \sim b \pmod{\mathfrak{m}}$, and $a \supset b$ implies $a \supset b \pmod{\mathfrak{m}}$. Hence the relations of division and equivalence modulo $\mathfrak{m}$ may be immediately extended to the residue classes $A, B, \cdots$ of $\mathfrak{D}$ modulo $\mathfrak{m}$.

THEOREM 4.2. *If $\mathfrak{m} = (m)$, then* (i) *$a \supseteq b \pmod{\mathfrak{m}}$ if and only if $(a, m) \supset b$ in $\mathfrak{D}$, and* (ii) *$a \sim b \pmod{\mathfrak{m}}$ if and only if $(a, m) \sim (b, m)$.*

Proof. (i) By theorem 3.1 (iii), $(a, m) \sim al + km$, l, m elements of $\mathfrak{D}$. Hence $(a, m) \supset b$ implies that $b = (al + km)q = aql + mqk$. Therefore by theorem 4.1, $(a, m) \supset b$ implies that $a \supset b \pmod{\mathfrak{m}}$. Next, if $a \supset b \pmod{\mathfrak{m}}$, then by theorem 4.1, $b = ar + ms$. Hence by theorem 3.1 (i), $(a, m) \supset b$.

(ii) If both $a \supset b \pmod{\mathfrak{m}}$ and $b \supset a \pmod{\mathfrak{m}}$, then $(a, m) \supset b$ and $(b, m) \supset a$. Hence $(a, m) \supset (b, m)$ and $(b, m) \supset (a, m)$ or $(a, m) \sim (b, m)$. The converse is evident.

THEOREM 4.3. $a \sim (a, m) \pmod{\mathfrak{m}}$.

Proof. $(a, m) \sim ((a, m), m)$.

THEOREM 4.4. *The correspondence $x \to x^* \sim (x, m)$ is a lattice homomorphism of $\mathfrak{D}$.*

Proof. Assume that $a \to a^*$ and $b \to b^*$. Then

$$(a, b)^* \sim (a, b), m) \sim ((a, m), (b, m)) \sim (a^*, b^*).$$

Since $\mathfrak{D}$ is a distributive lattice we also have

$$[a, b]^* \sim ([a, b], m) \sim [(a, m), (b, m)] \sim [a^*, b^*],$$

We observe that the correspondence of theorem 4.4 maps the lattice $\mathfrak{D}$ onto the sublattice of divisors of m in $\mathfrak{D}$.

THEOREM 4.5. *If* $a \to a^*$ *and* $b \sim a \pmod{\mathfrak{m}}$, *then* $b^* \sim a^*$, *and conversely.*

Proof. $a \to a^*$ if and only if $a^* \sim (a, m)$ and $a \sim b \pmod{\mathfrak{m}}$ if and only if $(a, m) \sim (b, m)$.

THEOREM 4.6. *The residue classes* $A, B, \cdots$ *of* $\mathfrak{O}$ *modulo* $\mathfrak{m}$ *form a lattice with respect to the relation of division modulo* $\mathfrak{m}$ *which is isomorphic with the lattice of the divisors of* $\mathfrak{m}$ *in* $\mathfrak{O}$ *if equivalent classes modulo* $\mathfrak{m}$ *and equivalent divisors are not regarded as distinct.*

Proof. We assign to A the element $a^* \sim (a, m)$, a any element of A. (We may if we please make the correspondence entirely definite by replacing a^* by the ideal $\mathfrak{a} = ((a), (m)) = (a^*)$). Then since $a \equiv b \pmod{\mathfrak{m}}$ implies $a \sim b \pmod{\mathfrak{m}}$, a^* is, up to equivalence, independent of the particular element of A used in defining it. By theorem 4.5, $A \sim B \pmod{\mathfrak{m}}$ if and only if $a^* \sim b^*$ and by theorem 4.2, $A \supset B \pmod{\mathfrak{m}}$ if and only if $a^* \supset b^*$. The result then follows from theorem 4.4.

PASADENA, CALIFORNIA.

REFERENCES

Morgan Ward, 1. *Duke Mathematical Journal*, vol. 3 (1937), pp. 627-636; 2. *Annals of Mathematics*, vol. 39 (1938), pp. 558-568.

Morgan Ward-R. P. Dilworth, 1. *Proceedings of the National Academy of Sciences*, vol. 24 (1938), pp. 162-164; 2. *Transactions of the American Mathematical Society.*

van der Waerden, 1. *Moderne Algebra*, vol. 1 (Berlin, 1930).

NOTE ON THE GENERAL RATIONAL SOLUTION OF THE EQUATION $ax^2 - by^2 = z^3$.*

By MORGAN WARD.

---

1. In a recent paper in this journal, E. Fogels (Fogels 1) has utilized the elements of algebraic number theory to determine all rational solutions of the diophantine equation

$$ax^2 - by^2 = z^3, \qquad a, b \text{ rational}, \neq 0. \tag{1}$$

I show here that the underlying reason for the success of Fogel's method is an arithmetical relationship between the degrees of the left and right sides of (1). For consider the more general equation

$$a_0x^m + a_1x^{m-1}y + \cdots + a_my^m = z^n \tag{2}$$

where m, n are positive integers and $a_0, \cdots, a_m$ rational. The general rational solution of (2) may be immediately written down provided that we impose the following condition: *The degrees m and n are co-prime.*

In this case we may reduce (2) to the simple equation

$$Y^m = Z^n. \tag{3}$$

2. There are four types of rational solutions of (2) to consider; namely, (i) solutions with $z = 0$, (ii) solutions with $x = 0$, (iii) solutions with $y = 0$ and (iv) solutions with none of x, y, z zero.

The first three types are readily treated.

(i) If $z = 0$, the solution of (2) may be expressed in the form $x = tx'$, $y = ty'$. Here t is an arbitrary rational and x', y' are either both zero or co-prime integers satisfying $a_0x'^m + \cdots + a_my'^m = 0$. Thus only a finite number of choices for x' and y' are possible.

(ii) If $x = 0$ and $a_m = 0$, then y is arbitrary, z is zero. If $a_m \neq 0$, then

---

* Received August 31, 1938.

(2) reduces to $a_m y^m = z^n$. Since m and n are co-prime, we may determine integers k and l such that

$$1 + km = ln. \tag{4}$$

On letting $y = a^k{}_m Y$, $z = a^l{}_m Z$ (2) is reduced to (3).

(iii) If $y = 0$, (2) is either trivial or similarly reducible to (3).

(iv) Let x, y, z be a rational solution of (2) with $xyz \neq 0$. Then if we let $x = uy$, u is rational and not zero, and (2) becomes

$$y^m w = z^n, \quad w = a_0 u^m + a_1 u^{m-1} + \ldots + a_m \neq 0.$$

As in type (ii), we let $y = w^k Y$, $z = w^l Z$. Then Y, Z are rational and non-zero and

$$Y^m = Z^n. \tag{3}$$

3. Equation (3) is a very special case of a type of diophantine system which I have already discussed in this journal. (Ward [1]). To solve (3), we let

$$Y = \Pi p^{\alpha}, \qquad Z = \Pi p^{\beta}$$

where the products extends over all primes, and the α and β are integers having only a finite number of non-zero values. Then (3) yields the condition $m\alpha = n\beta$. Hence $\alpha = (n : m)y$, $\beta = (m : n)y$ where y is an integer, and $m : n$ denotes the residual $m/(m, n)$ of n with respect to m. (Ward [2]). Since m and n are co-prime, $n : m = n$, $m : n = m$. Thus

$$Y = v^n, \qquad Z = v^m$$

where v is a rational number. On combining these results, we find that *if x, y, z is any rational solution of* (2) *with $xyz \neq 0$, then there exist rationals u and v such that*

$$\begin{aligned} x &= u(a_0 u^m + a_1 u^{m-1} + \cdots + a_m)^k v^n, \\ y &= (a_0 u^m + a_1 u^{m-1} + \cdots + a_m)^k v^n, \\ z &= (a_0 u^m + a_1 u^{m-1} + \cdots + a_m)^l v^m. \end{aligned} \tag{5}$$

Here k and l are integers satisfying

$$1 + km = ln. \tag{4}$$

Conversely, if u and v are rational and $a_0u^m + \cdots + a_m \neq 0$, $uv \neq 0$, then (5) always gives a rational solution of (2) with $xyz \neq 0$.

4. In particular, let $m = 2$, $n = 3$, $a_0 = a$, $a_1 = 0$, $a_2 = -b$. (2) then reduces to (1), and we may take $k = l = 1$ in (3). The formulas (5) become

$$\begin{aligned} x &= u(au^2 - b)v^3 \\ y &= (au^2 - b)v^3 \\ z &= (au^2 - b)v^2 \end{aligned}$$

If $a \neq 0$, we can make the reversible rational substitution $u = s/a$, $v = at$. We thus obtain

$$\begin{aligned} x &= as(s^2 - ab)t^3 \\ y &= a^2(s^2 - ab)t^3 \\ z &= a(s - ab)t^2. \end{aligned}$$

These are the formulas for the solution of (1) obtained by Fogels.

PASADENA, CALIFORNIA.

REFERENCES

E. Fogels, 1. *American Journal of Mathematics*, vol. 60 (July, 1938), pp. 734-736.
M. Ward, 1. *American Journal of Mathematics*, vol. 55 (1933), pp. 67-76; 2. *American Journal of Mathematics*, vol. 59 (1937), pp. 921-926.

ON EXTENSORS AND A EUCLIDEAN BASIS FOR HIGHER ORDER SPACES.*

By H. V. Craig.

1. Introduction. In previous papers [1] the writer has regarded certain aggregates of curves embedded in a Euclidean space as constituting a metric space of higher order; i. e. a space such that the arc length is given by the integral of a function containing derivatives of the base variables of order greater than one. In the present paper we shall regard such aggregates as a generalization of the concept surface and shall attempt to make manifest the analogy existing between tensors and surfaces on the one hand and extensors and generalized spaces on the other. During the course of this work we shall extend the treatment of extensors given in 7 studying in particular the Christoffel symbols and allied topics. We shall assume that the reader is acquainted with the material given in the first five articles of 7.

2. Notation. The notation to be employed is essentially that described in 7. Thus derivatives with respect to the parameter t will be designated for the most part by means of enclosed superscripts and partial derivatives will be indicated by subscripts, specifically:

$$
\begin{aligned}
&x'^a = dx^a/dt; \quad x^{(a)a} = d^a x^a/dt^a; \quad F_{(a)a} = \partial F/\partial x^{(a)a};\\
&X_l{}^a = X^{(0)a}_{(0)\,l}; \quad X^{(a)a}_{(\lambda)\,l} = \partial x^{(a)a}/\partial x^{(\lambda)\,l};\\
&X_l{}^{a(a)} = d^a X_l{}^a/dt^a.
\end{aligned}
\tag{2.1}
$$

We shall at times abridge this notation by omitting the Latin index from a pair of associated Greek and Latin indices. For example,

$$g_{aa\,.\,\beta b} X^{(a)a}_{(\lambda)\,l} X^{(\beta)b}_{(\nu)n} \text{ would be written } g_{a\beta} X_\lambda{}^a X_\nu{}^\beta.$$

Letters at the beginning of the alphabet $(a, b, c, d, e;\ \alpha, \beta, \gamma, \delta, \epsilon)$ are to be associated with the x-coördinate system while the others except i will serve to designate the $\bar{x}$-system.

Our summation convention is threefold: (a) repeated lower case Latin indices i excepted indicate summations 1 to N, (b) repeated lower case Greek indices indicate summations 0 to M (unless the contrary is indicated by the presence of a summation sign giving the range), (c) capital indices are not

* Received April 5, 1938; Revised (exposition only) December 1, 1938.

[1] See papers 2, p. 130; 5, p. 460; 6, p. 844 of the bibliography given in H. V. Craig, "On tensors relative to the extended point transformation," *American Journal of Mathematics*, vol. 59 (1937), p. 774. This paper will be referred to as 7 and papers listed in its bibliography will be designated by their numbers.

to be summed. In the sections treating of the Christoffel symbols and absolute differentiation, we frequently replace α with A, γ with Γ, ρ with R etc. in order to suspend the summation convention. Unrepeated Latin and Greek indices are to have the ranges N and M, respectively, unless the contrary is indicated. The index i has the range 1 to L.

3. Exsurfaces. Let there be given an N-dimensional space B (the base space), a coördinate system x^a, and the totality of parameterized arcs $x^a = x^a(t)$, lying in a region R of B—the functions $x^a(t)$ being of class M. On this basis we may formulate the definition:

DEFINITION (3.1). *The set of $N(M+1)$ numbers $x^{(a)a}$ associated with each point of a curve of class M will be said to be the coördinates of an extended point, or more briefly—expoint.*

Further, let $y^1, \cdots, y^L$ $(L > N)$ be a rectangular Cartesian coördinate system of an L-dimensional Euclidean space E. Evidently, a set of L equations,

$$(3.1) \quad y^i = y^i(x^1, \cdots, x^N; x'^1, \cdots, x'^N; \cdots; x^{(M)1}, \cdots, x^{(M)N})$$

associates with each expoint $x^{(a)a}$ of B, a point of E, and with each curve in parametric form, of class M, a curve in E, which we shall call the "warped curve."

DEFINITION (3.2). *The totality of warped curves will be called an "exsurface."*

This is obviously an extension of the notion surface—a point y^i of the exsurface being determined by an expoint (instead of a point) of the base space B. As a particular example we may let B be a Euclidean plane and E a Euclidean 3-space and regard equation (3.1) as distorting or warping the plane base curves into space curves. *If M is zero the exsurface is a surface. It should be borne in mind however that the exsurface is not merely a collection of points y^i but rather the totality of warped arcs.* The points of an exsurface might conceivably fill the entire space E while not every curve of E would be an exsurface curve. Consider, for example, $y^1 = x^1$, $y^2 = x^2$, $y^3 = x'^2/x'^1$. Obviously, the curve $y^2 = (y^1)^2$, $y^3 = 0$ is not an exsurface curve.

We shall classify the equations (3.1) into two classes: (1) some y^i is *not* invariant in functional form under an admissible parameter transformation; (2) each y^i is invariant in functional form and the arguments of y^i satisfy the requirement $x'^a x'^a \neq 0$. A parameter transformation $t = t(T)$, $a \leq T \leq b$ is "admissible" if the function $t(T)$ is increasing, single-valued, and of class M on (a, b).

As a simple generalization of (3.1) we could allow the parameter t to occur explicitly, thus in abbreviated notation y^i would be given by

$$y^i = y^i(t, x, x', \cdots, x^{(M)}). \tag{3.2}$$

The resulting manifold would be a Synge space [2]; i. e. the arc length would be given by the integral of a function containing t explicitly.

4. The first fundamental form, Class 1. If C is a base arc of class $(M+1)$ then the associated warped arc C is of class one and its derivative of arc is given by

$$(ds/dt)^2 = \sum_{i=1}^{L} y'^i y'^i = \sum_{i=1}^{L} \partial y^i/\partial x^{(\alpha)a} \cdot \partial y^i/\partial x^{(\beta)b} \cdot x^{(\alpha+1)a} x^{(\beta+1)b}, \tag{4.1}$$

which upon adopting the definition

$$g_{\alpha a . \beta b} = \sum_{i=1}^{L} \partial y^i/\partial x^{(\alpha)a} \cdot \partial y^i/\partial x^{(\beta)b} \tag{4.2}$$

may be written in the convenient form

$$(ds/dt)^2 = g_{\alpha a . \beta b} x^{(\alpha+1)a} x^{(\beta+1)b}. \tag{4.3}$$

We restrict ourselves to transformations of the base variables only and regard the various points and arcs as geometric invariants and functions such as y^i and y'^i as scalars. Accordingly, for each value of i, $\partial y^i/\partial x^{(\alpha)a}$ is an extensor of order one and therefore the quantities $g_{\alpha a . \beta b}$ constitute an extensor of order two.

The right member of (4.3), although it may be a *quadratic form in appearance only*, is an obvious generalization of the fundamental quadratic form of classical differential geometry. Accordingly, we introduce the definition:

DEFINITION (4.1). *The extensor $g_{\alpha a . \beta b}$ and the expression*

$$g_{\alpha a . \beta b} x^{(\alpha+1)a} x^{(\beta+1)b}$$

will be referred to as the metric extensor and the fundamental quadratic form, respectively.

We have pointed out in 7 that the quantities $x^{(\alpha+1)a}$ (α: 0 to M) are the components of an extensor—the tangent exvector. Consequently, we may sum up our results as follows:

THEOREM (4.1). *The square of the derivative of arc $(ds/dt)^2$ is the complete contraction of the metric extensor and the tangent exvector; i. e. it is the square of the "magnitude" of the tangent exvector.*

Evidently, we can form a determinant G from the quantities $g_{\alpha a . \beta b}$ by assigning the element $g_{\alpha a . \beta b}$ to the "αa" row and the "βb" column, the first row being denoted by 01, the second by 02, the N-th by $0N$, the $N+1$-st by 11 etc. Thus if $M = 2$, $N = 2$, we have

[2] See J. L. Synge (1).

$$(4.4) \qquad G = \begin{vmatrix} g_{01.01} & g_{01.02} & g_{01.11} & g_{01.12} & g_{01.21} & g_{01.22} \\ g_{02.01} & g_{02.02} & g_{02.11} & g_{02.12} & g_{02.21} & g_{02.22} \\ g_{11.01} & g_{11.02} & g_{11.11} & g_{11.12} & g_{11.21} & g_{11.22} \\ g_{12.01} & g_{12.02} & g_{12.11} & g_{12.12} & g_{12.21} & g_{12.22} \\ g_{21.01} & g_{21.02} & g_{21.11} & g_{21.12} & g_{21.21} & g_{21.22} \\ g_{22.01} & g_{22.02} & g_{22.11} & g_{22.12} & g_{22.21} & g_{22.22} \end{vmatrix}.$$

We note in passing that *the elements in the lower right hand* 2×2 *minor are tensor components while those in the similarly situated* 4×4 *minor constitute the components of an extensor of reduced range.*

We observe that if $L = (M+1)N$, G is the square of the Jacobian determinant $|\partial y^i/\partial x^{(a)a}|$, while if $L < (M+1)N$, G is always zero. Furthermore, by a well known theorem of matrix algebra, if $L > (M+1)N$ then G is the sum of the squares of the determinants of order $(M+1)N$ in the Jacobian matrix. Consequently, to insure the existence of the normalized cofactors $g^{aa.\beta b}$ we shall require that $L \geq (M+1)N$ and that J be of rank $(M+1)N$. If in addition $(M+1)N$ of the equations (3.1) admit of a unique single valued solution for the $x^{(a)}$'s in terms of the y's then we may assert that two different expoints do not determine the same exsurface point.

A second expression of a metric nature involving the extensor $g_{aa.\beta b}$ is furnished by the scalar product of two exsurface vectors. Thus, let C_1 and C_2 be two base curves defined by the equations: $x_1{}^a = x_1{}^a(t)$, $x_2{}^a = x_2{}^a(t)$ and further let us suppose that these equations are such that: $x_1{}^a(t_1) = x_2{}^a(t_2)$; $x'_1{}^a(t_1) = x'_2{}^a(t_2), \cdots, x_1{}^{(M)a}(t_1) = x_2{}^{(M)a}(t_2)$. Since these two base curves pass through the same expoint the corresponding warped curves $\tilde{C}_1$ and $\tilde{C}_2$ likewise intersect and the scalar product of the vectors dy_1/dt, dy_2/dt at this point of intersection is given by

$$(4.5) \qquad dy_1/dt \cdot dy_2/dt = \sum_{i=1}^{L} dy_1{}^i/dt \cdot dy_2{}^i/dt$$
$$= \sum_{i=1}^{L} (\partial y_1{}^i/\partial x^{(a)a} \cdot x_1{}^{(a+1)a})(\partial y_2{}^i/\partial x^{(\beta)b} \cdot x_2{}^{(\beta+1)b}).$$

Evidently, since the base curves C_1 and C_2 have an expoint in common at their point of intersection, the value of $\partial y^i/\partial x^{(a)a}$ at this intersection point is the same for both curves. Consequently, the metric extensors associated with these two base curves are identical at the point in question and the scalar product may be expressed as the contraction of the metric extensors with the two tangent exvectors, thus

$$(4.6) \qquad dy_1/dt \cdot dy_2/dt = g_{aa.\beta b} x_1{}^{(a+1)a} x_2{}^{(\beta+1)b}$$

or in abbreviated notation:

$$dy_1/dt \cdot dy_2/dt = g_{\alpha\beta} x_1^{(\alpha+1)} x_2^{(\beta+1)}. \tag{4.7}$$

To recapitulate we state the following theorem:

THEOREM (4.2). *If two base curves have an expoint P in common then the scalar product of the tangent vectors to the corresponding warped curves at their point of intersection is given by the contraction of the metric extensor and the two tangent exvectors at P.*

5. Exsurface normals. If a vector η is associated with the point $P[y^i(x^{(a)a})]$ of our exsurface and τ is a tangent vector to one of the warped curves through P then for a properly chosen number c

$$c\tau^i = dy^i/dt = \partial y^i/\partial x^{(a)a} \cdot x^{(a+1)a} \tag{5.1}$$

and consequently η is normal to τ if

$$\sum_{i=1}^{L} \eta^i \partial y^i/\partial x^{(a)a} \cdot x^{(a+1)a} = 0. \tag{5.2}$$

Evidently, for $L = (M+1)N + 1$ and each y^i linear in $x^{(a)a}$, $\alpha > 0$, x some coördinate system of our group, the equations

$$\sum_{i=1}^{L} \eta^i \partial y^i/\partial x^{(a)a} = 0 \tag{5.3}$$

admit of a single infinity of solutions which define a unique direction at P and furthermore although P may depend on the curve through $x^{(0)a}$ this direction does not. Obviously, there are exactly two unit vectors (oppositely directed) that satisfy (5.3). These unit vectors are normal to all warped curves whose base counterparts pass through the point $x^{(0)a}$, at the corresponding exsurface points P. Accordingly, we shall say that the vectors η *are normal to the exsurface in the strong sense.*

However, if the y^i are not all linear in $x^{(a)a}$, $\alpha > 0$ then the quantities η^i may vary with the base curve through $x^{(0)a}$. If they do, we confine ourselves to the collection of base curves passing through a given expoint $x^{(a)a}$, $\alpha = 0, \cdots, M$ and the condition for orthogonality becomes

$$\sum_{i=1}^{L} \eta^i \partial y^i/\partial x^{(M)a} = 0 \text{ and} \tag{5.4}$$

$$\sum_{i=1}^{L} \eta^i \Big(\sum_{a=0}^{M=1} \partial y^i/\partial x^{(a)a} \cdot x^{(a+1)a}\Big) = 0. \tag{5.5}$$

Obviously, (5.4) is a tensor equation and (5.4), (5.5) together imply the vanishing of $\Sigma \eta^i y'^i$, an invariant. Likewise (5.4) and $\Sigma \eta^i y'^i = 0$ imply (5.5). Therefore, we conclude that if (5.4) and (5.5) hold in one coördinate system x of our group, they are valid in all. A vector satisfying the foregoing requirements will be said to be *normal in the weak sense.*

For $L = (M+1)N + l$ and each y^i linear in $x^{(\alpha)a}$, $\alpha = 1, \cdots, M$, there exists an l-fold infinity of vectors η and accordingly there is determined a normal Euclidean l-space associated with the base point x^a.

7. Special exsurfaces, class 1. The exsurfaces of class 1 may be divided into two types: $L > (M+1)N$ and $L = (M+1)N$. Type 1 is characterized by the fact that the y's are functionally related and consequently the warped curves are a subset of the arcs of a Riemannian subspace of the manifold y^i. While in type 2 the y's are independent and the unit normal does not exist in the strong sense.

An important example occurring under both of these types is the explane. It may be defined as follows:

DEFINITION (7.1): *If for some coördinate system x in the base space an exsurface is defined by equations of the type*

$$y^i = C^i{}_{(\alpha)a} \cdot x^{(\alpha)a} + C^i \tag{7.1}$$

(the C's are constants) *then the exsurface will be said to be an explane of order M, M being the order of the derivative of highest order occurring in* (7.1).

If the explane is of type 1 its normals η are independent of the base point with which they are associated. Evidently the explanes (7.1) and

$$y^i = C^i{}_{(\alpha)a} \cdot x^{(\alpha)a} + b^i, \tag{7.2}$$

have common normal directions; consequently they will be said to be *parallel.*

Also worthy of mention are those exsurfaces which are such that for some coördinate system x one or more of the y's are functions of x alone. Obviously, a necessary and sufficient condition that y^i depend on x only is the vanishing of the extensor of reduced range $\partial y^i \partial x^{(\alpha)a}$, $\alpha > 0$. The tensor equivalent of this condition is ${}^vO_a y^i = 0$.[3]

Of considerable interest to us is the class of exsurfaces possessing a normal. Among these are those whose characteristic property is the existence of a coördinate system x in which the defining equations assume the form

$$y^i = f^i(x) + C^i{}_{(\alpha)\hat{a}} \cdot x^{(\alpha+1)a}. \tag{7.3}$$

Expanding in Taylor's series about the point x_0 (x_0 represents $x_0{}^1, \cdots, x_0{}^N$) and denoting by P_{y^i} the coördinates of a point of the explane,

$$P_{y^i} = f(x_0) + \partial f(x_0)/\partial x^a \cdot (x^a - x_0{}^a) + C^i{}_{(\alpha)a} x^{(\alpha+1)a} \tag{7.4}$$

we have

$$y^i = P_{y^i} + \text{higher degree terms.} \tag{7.5}$$

Evidently, the explane P_{y^i} has for its normal the normal to y^i at x_0 and furthermore if C is any base curve through x_0 and $\bar{C}$ and $\bar{C}(P)$ the correspond-

[3] See 6, p. 833.

ing exsurface and explane curves then $\tilde{C}$ and $\tilde{C}(P)$ are tangent. Consequently, we shall refer to P_{y^1} as the tangent explane "at x_0."

8. The Christoffel symbols. In this section we find it quite convenient to employ the abridgment described in the article on notation and accordingly we denote the Christoffel symbol $[\alpha a \beta b, \gamma c]$ by $[\alpha\beta, \gamma]$. To obtain its transformation equation we proceed as in ordinary tensor analysis. Thus denoting partial differentiation of the g's with respect to $x^{\gamma(c)}$ by $;\gamma$, we obtain upon differentiating

$$g_{\lambda\rho} = g_{a\beta} X_\lambda{}^a X_\rho{}^\beta \tag{8.1}$$

the equation

$$g_{\lambda\rho\,;\sigma} = g_{a\beta\,;\gamma} X_\lambda{}^a X_\rho{}^\beta X_\sigma{}^\gamma + g_{a\beta}[X_{\lambda\sigma}{}^a X_\rho{}^\beta + X_\lambda{}^a X_{\rho\sigma}{}^\beta]. \tag{8.2}$$

Following the well known procedure of permutation of indices, etc., we reach

$$[\lambda\rho, \sigma] = [\alpha\beta, \gamma] X_\lambda{}^a X_\rho{}^\beta X_\sigma{}^\gamma + g_{a\beta} X_{\lambda\rho}{}^a X_\sigma{}^\beta. \tag{8.3}$$

Multiplying, (8.3) by $X_\delta{}^\sigma$ and summing, we have

$$X_\delta{}^\sigma [\lambda\rho, \sigma] = [\alpha\beta, \delta] X_\lambda{}^a X_\rho{}^\beta + g_{a\delta} X_{\lambda\rho}{}^a. \tag{8.4}$$

Obviously, the normalized cofactors of the $g_{a\beta}$'s satisfy the relation

$$g^{\sigma\tau} X_\tau{}^\epsilon = g^{\delta\epsilon} X_\delta{}^\sigma. \tag{8.5}$$

Consequently, if we adopt the definition,

$$g^{\delta\epsilon}[\alpha\beta, \delta] = \left\{ \begin{matrix} \epsilon \\ \alpha\beta \end{matrix} \right\} \tag{8.6}$$

we may write

$$X_\tau{}^\epsilon \left\{ \begin{matrix} \tau \\ \lambda\rho \end{matrix} \right\} = \left\{ \begin{matrix} \epsilon \\ \alpha\beta \end{matrix} \right\} X_\lambda{}^a X_\rho{}^\beta + X_{\lambda\rho}{}^\epsilon. \tag{8.7}$$

Upon examining (8.3), (8.4) and (8.7), we observe at once a number of facts that are perhaps worthy of note. Thus, designating the Christoffel index which follows the comma as the *key index*, we see that *the Christoffel symbols (first kind) in the $\bar{x}$ system having key indices $\geq R$ are independent of the Christoffel symbols in the x-system having key indices $< R$.* Likewise, *the barred Christoffel symbols (of either kind) for which the sum of the symmetric indices $\geq R$ are independent of the unbarred symbols having the symmetric sum $< R$.* Accordingly, we make the following definition:

DEFINITION (8.1). *The set of Christoffel symbols of the first kind having key indices equal to or greater than R and symmetric sum equal to or greater than S $(R + S > 0)$ will be called* **Christoffel symbols of the first kind of reduced range.** Likewise, *the collection of symbols of the second kind having key index (superscript) equal to or less than R and symmetric sum equal to or greater than S $(M - R + S > 0)$ will be said to be of* **reduced range.**

Furthermore, by virtue of the fact that $X_\lambda{}^\alpha$ reduces to $X_l{}^a$ when $\alpha = \lambda = M$ and vanishes when λ exceeds α, while $X_{\lambda\rho}{}^\alpha$ vanishes whenever the sum of the subscripts exceeds the superscript, we may assert the theorem which follows.

THEOREM (8.1). *The completely reduced Christoffel symbols of the first kind* (i. e. *those having each Greek index equal to* M) *are the components of a* tensor, *while those of either kind having symmetric sum greater than* M *are the components of an* **extensor.** Also, *if* R *is any number of the set* $1, 2, \cdots, M$ *the Christoffel symbols of the second kind having key index less than* R *and symmetric sum equal to or greater than* R *are* **extensor components.**

As in Riemannian geometry, whenever the g's are constants in a certain coördinate system the Christoffel symbols in this system vanish and conversely. While if the g's are functions of x alone then all of the x-symbols of the first kind having each index $\leqq 1$ vanish and consequently all of the symbols (first kind) in each coördinate system having symmetric sum $> M$ and key index $\leqq 1$ vanish. Generalizing, we have the following theorem:

THEOREM (8.2). *Whenever the* g*'s in the* x*-system are functions of* $x, x', \cdots, x^{(R)}$ *only, the set of Christoffel symbols of the first kind having each index greater than* R *and symmetric sum greater than* M *vanishes in all coördinate systems.*

With regard to the range of the g's in the last term of (8.3) we note that if $\lambda + \rho = \sigma = M$, this term reduces to $g_{M_a M_b} X^{M_a}_{\lambda l \rho r} X_s{}^b$. But by virtue of the formula

$$X_\Lambda{}^A = \binom{A}{\Lambda} X_l{}^{a(A-\Lambda)} = \binom{A}{\Lambda} X_l{}^{(A-\Lambda)a} \tag{8.8}$$

it follows that

$$X_{\Lambda R}{}^A = \binom{A}{\Lambda} X^{(A-\Lambda)a}_{Rr.l} = \binom{A}{\Lambda}\binom{A-\Lambda}{R} X^{a(A-\Lambda-R)}_{rl} . \tag{8.9}$$

Consequently, $X_{\lambda\rho}{}^M (\lambda + \rho = M)$ may be written in the form $\binom{M}{\lambda} X_{rl}{}^a$. Now if the determinant $|\, g_{M_a M_b} \,|$ (which occurs in the lower right-hand corner of the matrix) does not vanish, we can form a connection which will serve as a means of eliminating the second derivative $X_{rl}{}^a$. Thus, if the symbols $g_0{}^{ab}$ represent the normalized cofactors in the lower right $N \times N$ determinant and we restrict $\lambda + \rho$ and σ in (8.3) to the value M, then we may write successively,

$$[\lambda\rho, Ms] = [\alpha\beta, Mc] X_\lambda{}^\alpha X_\rho{}^\beta X_s{}^c + \binom{M}{\lambda} g_{M_a M_b} X_{rl}{}^a X_s{}^b \tag{8.10}$$

and

$$X_t{}^d \left\{ \begin{matrix} t \\ \lambda\rho \end{matrix} \right\}^0 = \left\{ \begin{matrix} d \\ \alpha\beta \end{matrix} \right\}^0 X_\lambda{}^\alpha X_\rho{}^\beta + \binom{M}{\lambda} X_{rl}{}^d . \tag{8.11}$$

The definition of the new symbols of the second kind being

$$\left\{ \begin{matrix} d \\ \alpha\beta \end{matrix} \right\} = g_0{}^{ed}\,[\alpha\beta, Me] \tag{8.12}$$

and their transformation equation for $\lambda + \rho > M$ being (8.11) with the last term deleted. Evidently, the connection (8.11) differs from that obtained by substituting $\epsilon = M, \lambda + \rho = M$ into (8.7).

9. Excovariant differentiation. Following a well known procedure of tensor analysis, differentiation processes may be obtained from either (8.7) or (8.11). The derivatives so obtained differ from those of tensor analysis in one respect at least; namely, some of the components of the derivatives are merely a sum or difference of extensors. To illustrate, let V_{aa} be a differentiable extensor and let us, as before, denote partial derivatives by means of subscripts preceded by a semicolon. Thus, in abridged notation, we have

$$V_\lambda = V_a X_\lambda{}^a \tag{9.1}$$

and

$$V_{\lambda;\rho} = V_{a;\beta} X_\lambda{}^a X_\rho{}^\beta + V_\epsilon X_{\lambda\rho}{}^\epsilon. \tag{9.2}$$

While multiplication of (8.7) by V_ϵ yields

$$V_\tau \left\{ \begin{matrix} \tau \\ \lambda\rho \end{matrix} \right\} = V_\epsilon \left\{ \begin{matrix} \epsilon \\ \alpha B \end{matrix} \right\} X_\lambda{}^a X_\rho{}^\beta + V_\epsilon X_{\lambda\rho}{}^\epsilon \tag{9.3}$$

and consequently the excovariant derivative $V_{a,\beta}$ defined by

$$V_{a,\beta} = V_{a;\beta} - V_\epsilon \left\{ \begin{matrix} \epsilon \\ \alpha\beta \end{matrix} \right\}, \tag{9.4}$$

is an extensor of order two. We note that for $\lambda + \rho > M$ *equations* (9.2) and (9.3) *are merely extensor transformations of reduced range.*

Similarly, the excovariant derivative of an excontravariant vector V^ϵ is given by

$$V^\epsilon{}_{,\beta} = V^\epsilon{}_{;\beta} + \left\{ \begin{matrix} \epsilon \\ \alpha\beta \end{matrix} \right\} V^a \tag{9.5}$$

and the components having $\epsilon < \beta$ are the sum of extensors. The formal equivalence of these derivatives and those of ordinary tensor analysis leads us to assert that the theorems relating to the covariant derivatives of sums, products and contractions have counterparts with respect to extensors.

Obviously, *the analogue of Ricci's lemma*

$$g_{\alpha\beta,\gamma} = 0 \tag{9.6}$$

holds. Furthermore, *the product of Kronecker deltas* $\delta_\beta{}^a \delta_b{}^a$ *is an extensor of second order and*

$$(9.7)\qquad (\delta_\beta{}^a\delta_b{}^a)_{,\gamma c} = \delta_\beta{}^\epsilon\delta_b{}^e \left\{ \begin{matrix} \alpha a \\ \epsilon e \cdot \gamma c \end{matrix} \right\} - \delta_\epsilon{}^a\delta_e{}^a \left\{ \begin{matrix} \epsilon e \\ \beta b \cdot \gamma c \end{matrix} \right\} = 0.$$

Consequently, since

$$(9.8)\qquad g^{\alpha a \cdot \beta b} g_{\alpha a \cdot \gamma c} = \delta_\gamma{}^\beta \delta_c{}^b,$$

we have

$$(9.9)\qquad g^{\alpha a \cdot \beta b},\quad \delta d = 0.$$

It should be observed, however, that while the complete contraction $T^a{}_{a,\gamma}$ of the excovariant derivative $T^a{}_{\beta,\gamma}$ reduces to the partial derivative of the invariant $T_a{}^a$, the corresponding reduction does not take place in the case of contraction over a reduced range. Thus, contracting the derivative $T^\alpha{}_{\beta,\gamma}$,

$$(9.10)\qquad T^\alpha{}_{\beta,\gamma} = T^\alpha{}_{\beta;\gamma} + T_\beta{}^\delta \left\{ \begin{matrix} \alpha \\ \delta\gamma \end{matrix} \right\} - T_\delta{}^\alpha \left\{ \begin{matrix} \delta \\ \beta\gamma \end{matrix} \right\}$$

we get

$$(9.11)\qquad \sum_{\beta=E}^{M} \binom{\beta}{E} T^{\beta-E}_{\beta,\gamma} = \sum_{\beta=E}^{M} \binom{\beta}{E} \left[T^{\beta-E}_{\beta;\gamma} + T_\beta{}^\delta \left\{ \begin{matrix} \beta - E \\ \delta \quad \gamma \end{matrix} \right\} - T_\delta{}^{\beta-E} \left\{ \begin{matrix} \delta \\ \beta\gamma \end{matrix} \right\} \right].$$

To recapitulate, we state the theorem which follows.

THEOREM (9.1). *The process defined by the equation*

$$T^{\alpha\ldots}_{\beta\ldots,\gamma} = T^{\alpha\ldots}_{\beta\ldots;\gamma} + T^{\delta\ldots}_{\beta\ldots} \left\{ \begin{matrix} \alpha \\ \delta\gamma \end{matrix} \right\} - T^{\alpha\ldots}_{\delta\ldots} \left\{ \begin{matrix} \delta \\ \beta\gamma \end{matrix} \right\}, \ldots,$$

is such that: (a) *when applied to an extensor of any order it yields a second of excovariant order one greater than that of the original;* (b) *it obeys the analogues of the theorems of tensor analysis relating to the covariant derivatives of sums, products and contractions* (full range) *but the analogy breaks down with regard to contractions over a reduced range; and* (c) *it yields a zero extensor when applied to the extensors:* $g_{\alpha a \cdot \beta b}$, $g^{\alpha a \cdot \beta b}$, and $\delta_\beta{}^a\delta_b{}^a$.

With regard to the second connection (8.11), we shall see that this relationship has applications to the differentiation of both tensors and extensors. As a first example, we develop a differentiation process for a vector V whose components are differentiable functions of $x, x', \cdots, x^{(M)}$. Thus differentiating

$$(9.12)\qquad V_r = V_b X_r{}^b$$

we have

$$(9.13)\qquad V_{r;l} = V_{b;a} X_r{}^b X_l{}^a + V_d X^d{}_{rl}.$$

While, substituting $\lambda = 0$ in (8.11) and multiplying the result by V_d leads to

$$(9.14)\qquad V_t \left\{ \begin{matrix} t \\ 0l \cdot Mr \end{matrix} \right\}^0 = V_d \left\{ \begin{matrix} d \\ \alpha Mb \end{matrix} \right\}^0 X_l{}^a X_r{}^b + V_d X^d{}_{rl},$$

consequently,

$$(9.15)\qquad V_{r;l} - V_t \left\{ \begin{matrix} t \\ 0lMr \end{matrix} \right\}^0 = \left[V_{b;a} - V_d \left\{ \begin{matrix} d \\ \alpha \cdot Mb \end{matrix} \right\}^0 \right] X_r{}^b X_l{}^a.$$

Now for $\lambda \neq 0$, we have

$$V_{r;\lambda} = V_{b;a} X_r{}^b X_\lambda{}^a \tag{9.16}$$

and

$$V_t \left\{ \begin{matrix} t \\ \lambda \cdot Mr \end{matrix} \right\}^0 = V_d \left\{ \begin{matrix} d \\ \alpha Mb \end{matrix} \right\}^0 X_r{}^b X_\lambda{}^a. \tag{9.17}$$

Consequently, $V_{b;a} - V_d \left\{ \begin{matrix} d \\ Mb \cdot \alpha \end{matrix} \right\}$ *is a mixed extensor (covariant, exco-variant).*

If on the other hand V is an exvector and we so restrict λ and ρ in (9.2) that $\lambda + \rho \geq M$ then the only second derivative involved is $X^e{}_{lr}$. Evidently, this derivative can be eliminated from the modified form of (8.11),

$$X_t{}^d \left\{ \begin{matrix} t \\ \lambda\rho \end{matrix} \right\}^0 = \left\{ \begin{matrix} d \\ \alpha\beta \end{matrix} \right\}^0 X_\lambda{}^a X_\rho{}^\beta + X^{(M)d}_{\lambda\rho}, \quad \lambda + \rho \geq M. \tag{9.18}$$

From this result we conclude that

$$V_{a;\beta} - V_{Md} \left\{ \begin{matrix} d \\ \alpha\beta \end{matrix} \right\}, \quad \alpha + \beta \geq M, \tag{9.19}$$

is an extensor of reduced range.

Finally, as an example of a simultaneous application of (8.7) and (8.11) let us consider the problem of constructing a differentiation process for the mixed extensor $T_a{}^\gamma$.

Differentiating the transformation equation

$$T_l{}^\sigma = T_a{}^\gamma X_\gamma{}^\sigma X_l{}^a \tag{9.20}$$

we obtain

$$T^\sigma{}_{l;\rho} = T^\gamma{}_{a;\beta} X_\rho{}^\beta X_\gamma{}^\sigma X_l{}^a + T_a{}^\delta X^\sigma{}_{\delta\beta} X_\rho{}^\beta X_l{}^a + T_d{}^\gamma X_\gamma{}^\sigma X^d{}_{lr} \delta_\rho{}^0. \tag{9.21}$$

By virtue of (8.7) we may replace the second term with the expression

$$T_a{}^\delta \left[X_\gamma{}^\sigma \left\{ \begin{matrix} \gamma \\ \beta\delta \end{matrix} \right\} - \left\{ \begin{matrix} \sigma \\ \mu\nu \end{matrix} \right\} X_\beta{}^\mu X_\delta{}^\nu \right] X_\rho{}^\beta X_l{}^a \tag{9.22}$$

and if $\rho = 0$ the last term may be written

$$T_d{}^\gamma \left[X_l{}^a \left\{ \begin{matrix} t \\ Ml \cdot 0r \end{matrix} \right\}^0 - \left\{ \begin{matrix} d \\ Ma \cdot \beta \end{matrix} \right\}^0 X_l{}^a X_r{}^\beta \right] X_\gamma{}^\sigma. \tag{9.23}$$

Thus for $\rho = 0$ the following holds:

$$\begin{aligned} T^\sigma{}_{l;\rho} + T_l{}^\nu \left\{ \begin{matrix} \sigma \\ \rho\nu \end{matrix} \right\} - T_l{}^\sigma \left\{ \begin{matrix} t \\ Ml \cdot \rho \end{matrix} \right\}^0 \\ = \left[T^\gamma{}_{a;\beta} + T_a{}^\delta \left\{ \begin{matrix} \gamma \\ \beta\delta \end{matrix} \right\} - T_d{}^\gamma \left\{ \begin{matrix} d \\ Ma \cdot \beta \end{matrix} \right\}^0 \right] X_\gamma{}^\sigma X_l{}^a X_\rho{}^\beta. \end{aligned} \tag{9.24}$$

But the expression (9.23) with $0r$ replaced by ρr vanishes for $\rho > 0$, consequently, this modified form of (9.23) may be substituted for the third term

of (9.21) regardless of the value of ρ. Therefore, equation (9.24 is valid without restriction, and we have the following theorem:

THEOREM (9.2). *If $T_a{}^\gamma$ is a differentiable mixed extensor (tensor-extensor) then*

$$T^\gamma{}_{a;\beta} + T_a{}^\delta \left\{ \begin{matrix} \gamma \\ \beta\delta \end{matrix} \right\} - T_d{}^\gamma \left\{ \begin{matrix} d \\ Ma \cdot \beta \end{matrix} \right\}^0$$

is a mixed extensor of the type indicated by the free indices γ, a, and β.

10. Miscellaneous remarks on absolute differentiation. Intrinsic derivatives may be formed formally just as in tensor analysis; however, the special character of our transformations (viewed from the $(M+1)N$ dimensional standpoint) leads us to certain conclusions not having a counterpart in ordinary tensor analysis.

For example, if $V^{\alpha a}$ is an extensor formed from a vector by differentiation, i. e. $V^{\alpha a} = V^{a(\alpha)}$, then an additional differentiation of V may be regarded as merely extending the range. Thus, the original quantities follow the transformation law,

$$V^a = V^\rho X_\rho{}^a; \qquad (\alpha, \rho = 0, 1, \cdots, M); \tag{10.1}$$

while the addition of the new quantities $V^{(M+1)a}$, $V^{(M+1)r}$ yields sets satisfying the similar relationship

$$V^a = V^\rho X_\rho{}^a; \qquad (\rho, \alpha = 0, 1, \cdots, M+1). \tag{10.2}$$

Consequently, in this case the need for a special differentiation process is not as urgent as in tensor analysis. The intrinsic derivative of V is, of course, given by the expression

$$DV^a = (V^a)' + V^\beta \left\{ \begin{matrix} \alpha \\ \beta\gamma \end{matrix} \right\} x'^\gamma \tag{10.3}$$

and has the advantage that the range of indices is unaltered.

It is interesting to observe that if $V^{\alpha a} = V^{a(\alpha)}$, $\alpha < \Omega$ then quantities of reduced range may be formed by subtraction. For example, since $V^{\alpha a} - V^{a(\alpha)}$, $\alpha < \Omega$, is an extensor of reduced range, if its components vanish in one coördinate system they do in all, and consequently we have, for $\alpha \gtrless M - \Omega$,

$$V^{\Omega+\alpha \cdot a} - V^{a(\Omega+\alpha)} = \sum_{\lambda=\Omega}^{M} [V^{\lambda l} - V^{l(\lambda)}] X_{\lambda l}^{\Omega+\alpha \cdot a}. \tag{10.4}$$

Multiplying by $A!/(\Omega+A)!$, replacing λ with $\lambda+\Omega$ and substituting from the formula

$$A!/(\Omega+A)! \cdot X_{\Omega+\Lambda}^{\Omega+A} = \Lambda!/(\Omega+\Lambda)! \cdot X_\Lambda{}^A, \tag{10.5}$$

we obtain

$$A!/(\Omega+A)! \cdot [V^{\Omega+A \cdot a} - V^{a(\Omega+A)}] = \sum_{\lambda=0}^{M-\Omega} [V^{\Omega+\lambda \cdot l} - V^{l(\Omega+\lambda)}] \lambda!/(\Omega+\lambda)! \cdot X_{(\lambda) l}^{(A) a}. \tag{10.6}$$

Hence we may assert the theorem:

THEOREM (10.1). *If* $V^{\alpha a} = V^{a(\alpha)}$, $\alpha < \Omega$, *then the quantities*

$$A!/(\Omega + A)! \cdot [V^{\Omega+A\,.\,a} - V^{a(\Omega+A)}], \qquad (A = 0, \cdots, M - \Omega),$$

are the components of an extensor of reduced range.

Thus it is possible to form extensors of reduced range by differentiation without the introduction of Christoffel symbols.

Usually, Ω is limited to the value unity, while M in this particular problem is limited only by the requirement $M \geqq \Omega$. For $M = \Omega$ (10.4) reduces to the *tensor transformation,*

$$V^{\Omega a} - V^{a(\Omega)} = [V^{\Omega l} - V^{l(\Omega)}]X_l{}^a. \tag{10.7}$$

Evidently, the intrinsic derivative may be obtained from the covariant by multiplying by $x^{(\beta+1)}$ and summing on β from zero to M. Further, since $V^a{}_\beta$ is an extensor, we may contract with $x^{(\beta+1)}$ over a reduced range. Thus, *the quantities*

$$\sum_{\beta=E}^{M} \binom{\beta}{E} V^a{}_{,\beta}\, x^{(\beta+1-E)} \tag{10.8}$$

are the components of an exvector. This generalization, however, does not exhaust the possibilities, for we shall see that the transformation

$$\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} X_\lambda{}^a = A x^{(A)}, \qquad A = \alpha, \tag{10.9}$$

leads to other absolute derivatives.

As a matter of fact, (10.9) is a special case of the theorem which follows.

THEOREM (10.2). *If* V^a *is any exvector and* Γ *is equal to or less than* Λ *which in turn is equal to or less than* M *then the transformation equation of* $V^{a\text{-}\Gamma}$, $\alpha = \Gamma, \cdots, M$, *may be put in the form*

$$\sum_{a=\Gamma}^{M} \binom{\alpha}{\Gamma} V^{a\text{-}\Gamma} X_a{}^\lambda = \binom{\Lambda}{\Gamma} V^{\Lambda\text{-}\Gamma}, \qquad (\Lambda \gtrless M). \tag{10.10}$$

Proof. As a first step we note that

$$\begin{aligned} \binom{A}{\Gamma} X_A{}^\lambda &= \binom{A}{\Gamma}\binom{\Lambda}{A}\binom{\Lambda - \Gamma}{A - \Gamma}^{-1} X^{\Lambda\text{-}\Gamma}_{A\text{-}\Gamma} \\ &= \binom{\Lambda}{\Gamma} X^{\Lambda\text{-}\Gamma}_{A\text{-}\Gamma}. \end{aligned} \tag{10.11}$$

Consequently, the left member of (10.10) may be put in the form

$$\sum_{a=\Gamma}^{M} \binom{\Lambda}{\Gamma} V^{(a\text{-}\Gamma)} X^{\Lambda\text{-}\Gamma}_{a\text{-}\Gamma} \tag{10.12}$$

which upon replacing α with $\alpha + \Gamma$ reduces to the right member and the

theorem is established. Evidently, this theorem admits of generalization to higher order quantities, such as, $\binom{A}{\Gamma}\binom{B}{\Delta} T^{A-\Gamma,B-\Delta}$.

Returning to (10.9), differentiating it with respect to x^ρ and x^σ, and replacing ρ and σ with the corresponding capitals R and Σ to prevent summation, we obtain successively

$$\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} X^A{}_{\lambda R} = (A - R) X_R{}^A \tag{10.13}$$

and

$$\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} X^A{}_{\lambda R\Sigma} = (A - R - \Sigma) X^A{}_{R\Sigma} \tag{10.14}$$

etc. Consequently, derivatives may be reduced by multiplication by $\lambda x^{(\lambda)}$.

A generalization of (10.13) may be obtained by first replacing V in (10.10) by x' and then differentiating with respect to $x^{(\beta)}$. The result is

$$\binom{\Gamma + B - 1}{\Gamma} X^\lambda{}_{\Gamma+B-1} + \sum_{\alpha=\Gamma}^{M} \binom{\alpha}{\Gamma} x^{(\alpha+1-\Gamma)} X^\lambda{}_{\alpha\beta} = \binom{\Lambda}{\Gamma} X_\beta{}^{\Lambda+1-\Gamma}, \tag{10.15}$$

which upon replacing

$$\binom{\Gamma + \beta - 1}{\Gamma} X^\lambda{}_{\Gamma+\beta-1} \quad \text{by} \binom{\Lambda}{\Gamma} \frac{\beta}{\Lambda - \Gamma + 1} X_\beta{}^{\Lambda+1-\Gamma}$$

assumes the form

$$\sum_{\alpha=\Gamma}^{M} \binom{\alpha}{\Gamma} x^{(\alpha+1-\Gamma)} X_{\alpha\beta}{}^\Lambda = \binom{\Lambda}{\Gamma} [1 - \beta/(\Lambda - \Gamma + 1)] X_\beta{}^{\Lambda+1-\Gamma}. \tag{10.16}$$

As an application of (10.9), we prove that the contraction $\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} T_\lambda$ is an invariant; thus

$$\begin{aligned} \sum_{\lambda=1}^{M} \lambda x^{(\lambda)} T_\lambda &= \sum_{\lambda=1}^{M} \sum_{a=1}^{M} T_a (\lambda x^\lambda X_\lambda{}^a) \\ &= \sum_{a=1}^{M} a x^{(a)} T_a. \end{aligned} \tag{10.17}$$

Evidently, equation (10.10) *and its generalization provide a natural introduction to contraction over a reduced range.*

As a second illustration of *reduction of a derivative* we multiply

$$T_{\lambda;\rho} = T_{a;\beta} X_\lambda{}^a X_\rho{}^\beta + T_a X^a{}_{\lambda\rho}; \qquad \lambda, \rho \geqq 1; \tag{10.18}$$

by $\lambda x^{(\lambda)}$ and sum from 1 to M. The result is

$$\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} T_{\lambda;R} = \Big(\sum_{a=1}^{M} a x^{(a)} T_{a;\beta} \Big) X_R{}^\beta + \sum_{a=1}^{M} (a - R) X_R{}^a T_a \tag{10.19}$$

from which we conclude that

$$\sum_{\lambda=1}^{M} \lambda x^{(\lambda)} T_{\lambda;\,R} + RT_R \tag{10.20}$$

is an extensor or reduced range.

11. Exsurfaces independent of the parameter. It will be recalled that we divided exsurfaces into two classes. The second of these, the class with which we are now concerned, was characterized by the restrictions: (a) each base arc is such that $\Sigma(dx^a/dt)^2$ is never zero; (b) each function $y^i(x, x', \cdots, x^{(M)})$ is invariant in functional form under an admissible parameter transformation. Denoting the result of substituting $t(T)$ for t in $x(t)$ by X and representing d^aX^a/dT^a by $X^{(a)a}$ the meaning of condition (b) may be stated as follows: if we replace t in $y(x^{(a)a})$ by $t(T)$ the result is $y(X^{(a)a})$, i. e., the functional relationship existing between $y(t)$ and its intermediate arguments x, $dx/dt, \cdots, d^Mx/dt^M$ is identical with the functional relationship holding between $y(T)$ and its arguments X, $dX/dT, \cdots, d^MX/dT^M$. Further, for corresponding values of t and T, $(t = t(T))$ we have $y(x^{(a)a}) = y(X^{(a)a})$. Consequently, the warped curve (geometrical locus) associated with a given base arc $x = x(t)$ is invariant under a change of parameter, while in the case of class one the parameter had an intrinsic meaning (time for example) and parameter transformations were implicitly ruled out, each expoint having a unique set of $(M+1)N$ coördinates. In the present case however, *two sets of numbers such as $x^{(a)a}, X^{(a)a}$ will be regarded as the coördinates of the same base expoint if there exists an admissible parameter transformation which will convert one set into the other.* If we apply such a transformation to $x(t)$ we get $x(t) = x[t(T)]$; and denoting $x[t(T)]$ by $X(T)$ we have, identically in t, $x(t) = X[T(t)]$. Differentiation with respect to t yields the identity $x' = X'dT/dt$, $(x' = dx/dt, X' = dX/dT, dT/dt > 0)$. Thus we see that *for the case $M = 1$, two sets x^a, x'^a; X^a, X'^a are the coördinates of the same expoint if and only if $x^a = X^a$ while $x'^a = kX'^a$,* $k > 0$, $\Sigma(x'^a)^2 \neq 0$. We note in passing that the prime coördinates may be regarded as homogeneous or projective coördinates. To continue, *if $M = 2$, we have, as the condition for equivalence of $x^{(a)a}$ and $X^{(a)a}$, the following relationships:* $x^a = X^a$, $x'^a = X'^a k$; $x''^a = X''^a k^2 + X'^a l$. $(l = d^2T/dt^2)$; $\Sigma(x'^a)^2 \neq 0$. Now one of the numbers x'^a is not zero and consequently k and l are never indeterminate. Thus if x'^1 is not zero k and l are given by

$$k = x'^1/X'^1;\ l = [x''^1 - X''^1(x'^1/X'^1)^2]/X'^1. \tag{11.1}$$

Equivalence tests for the cases $M > 2$ may be obtained similarly.

In addition to securing the invariance under parameter transformation of the warped curve associated with a given base arc, it should be noted that, *condition* (b) *implies the invariance in functional form of ds regarded as a function of $x^{(a)a}$ and dt,* for

$$(11.2)\qquad \frac{dy(l)}{dt}=\frac{\partial y(l)}{\partial x^{(a)a}}\cdot x^{(a+1)a}=\frac{dy(T)}{dT}\ \frac{dT}{dt}=\frac{\partial y(T)}{\partial X^{(a)a}}\cdot X^{(a+1)a}\frac{dT}{dt}.$$

A third and perhaps unfortunate consequence of (b) *is that it implies the vanishing of the determinant of the metric extensor.* This may be inferred from the fact that a continuum of ordered number sets $x^{(a)a}$, (all sets coinciding in the first N numbers x^a) may determine but a single point y^i. To examine this matter further we recall that (b) implies the conditions [4]

$$(11.3)\qquad \sum_{\delta=A}^{M}\binom{\delta}{A}x^{(\delta-A+1)b}\cdot(yH)_{(\delta)b}=\delta_A{}^1yH,\qquad A\geqq 1,$$

for each function y of our set and H any function of x, x' positively homogeneous of degree plus one in x', since $yHdt$ is invariant. From (11.3) and the fact that $x'^aH_{(1)a}=H$, we derive

$$(11.4)\qquad \sum_{\delta=A}^{M}\binom{\delta}{A}x^{(\delta-A+1)b}y_{(\delta)b}=0,\qquad A\geqq 1.$$

By virtue of the foregoing relationship we can transform the Jacobian matrix,

$$(11.5)\qquad \begin{pmatrix} y^1{}_{(0)a}y^1{}_{(1)a} & \cdots & y^1{}_{(M)a}\\ \cdot & & \\ \cdot & & \\ \cdot & & \\ y^L{}_{(0)a}y^L{}_{(1)a} & \cdots & y^L{}_{(M)a}\end{pmatrix}$$

without altering its rank so that each set of N columns after the first set of N will contain exactly one column of zeros. Hence we have

THEOREM (11.1). *It is impossible to have a set of* $(M+1)N$, $M>0$, *functionally independent functions* $y(x^{(a)a})$ *each invariant in functional form under a parameter transformation.*

An immediate consequence of this theorem and the relationship between the determinant $g_{a\alpha\,.\,\beta b}$ and the Jacobian, is the following:

COROLLARY. *The determinant of the metric extensor of an exsurface of class two and order* $M>0$ *vanishes.*

A fourth consequence of (b) is

THEOREM (11.2). *The components of the metric extensor of an exsurface of class two and order* $M>0$ *cannot be functions of* x *alone.*

Proof. If, as in Finsler geometry, we denote $(ds/dt)^2$ by $\frac{1}{2}F^2$ or f then Fdt is invariant in functional form under a parameter transformation and consequently we may substitute F for yH in (11.3) provided we replace M

[4] See, H. V. Craig, 5, p. 461.

with $M+1$. Thus, if we suppress the Latin indices and denote $\partial F/\partial x^{(\delta)}$ by F_δ, we have

$$\sum_{\delta=\Gamma}^{M+1}\binom{\delta}{\Gamma}x^{(\delta-\Gamma+1)}F_\delta=\delta_\Gamma{}^1F,\ \Gamma\geq 1, \tag{11.6}$$

which, after multiplying by F, takes the form

$$\sum_{\delta=\Gamma}^{M+1}\binom{\delta}{\Gamma}x^{(\delta-\Gamma+1)}f_\delta=2\delta_\Gamma{}^1f,\ \Gamma\geq 1. \tag{11.7}$$

Assuming that the g's are functions of x alone, we may write

$$f_\delta=2x^{(\alpha+1)}g_{\alpha\,.\,\delta-1},\qquad(\delta=1,\cdots,M+1), \tag{11.8}$$

which upon substitution in (11.7) with $\Gamma=1$, yields

$$\sum_{\delta=1}^{M+1}2\delta x^{(\delta)}x^{(\alpha+1)}g_{\alpha\,.\,\delta-1}=2f \tag{11.9}$$

or replacing δ with $\delta+1$

$$2\delta x^{(\delta+1)}x^{(\alpha+1)}g_{\alpha\delta}=0;\qquad(\alpha,\delta=0,1,\cdots,M). \tag{11.10}$$

Hence we conclude that the $g_{\alpha\delta}$ are all zero—a contradiction.

We note in passing that *the equations* (11.6) *and* (11.7) *are generalizations of certain fundamental identities*[5] *of Finsler geometry. Other generalizations or analogues may be obtained by differentiation, multiplication, and contraction. Thus,* indicating differentiation by subscripts alone, *we have*

$$\sum_{\alpha=1}^{M+1}\alpha x^{(\alpha)}F_{\alpha\Gamma}=(1-\Gamma)F\Gamma,\qquad\Gamma\geq 1, \tag{11.11}$$

$$\sum_{\alpha=1}^{M+1}\alpha x^{(\alpha)}F_{\alpha\Gamma\Delta}=(1-\Gamma-\Delta)F\Gamma\Delta,\qquad\Delta,\Gamma\geq 1, \tag{11.12}$$

$$\sum_{\alpha=1}^{M+1}\alpha x^{(\alpha)}f_{\alpha\Delta}=(2-\Delta)f\Delta,\qquad\Delta\geq 1, \tag{11.13}$$

$$\sum_{\alpha=1}^{M+1}\sum_{\beta=1}^{M+1}\alpha x^{(\alpha)}x^{(\beta)}[-FF_{\alpha\beta}+F_\alpha B_\beta]$$
$$=\sum_{\beta=1}^{M+1}(B-1)x^{(\beta)}FF_\beta+x^\beta FF_\beta=F^2. \tag{11.14}$$

With regard to the special exsurfaces of the second class, we note that the explane as defined in article 7 is ruled out by condition (b). The following locus, which we shall call a pseudo plane, however, may serve as a substitute:

$$y^1=x^1;\ y^2=x^2;\ y^3=Ax^1+Bx^2+C, \tag{11.15}$$

the quantities A, B, and C being functions of the slope m, $(m=x'^2/x'^1)$.

---

[5] See J. H. Taylor, "A generalization of Levi-Civita's parallelism and the Frenet formulas," *Transactions of the American Mathematical Society*, vol. 27 (1925), p. 248.

Evidently, (11.15) warps the set of all base lines of slope m into a set of parallel lines. However, the pseudo plane is not built up entirely of plane curves. Two pseudo planes, $(A, B. C; \bar{A}, \bar{B}, \bar{C},)$ will be said to be parallel if $A = \bar{A}$, $B = \bar{B}$, and $C = \bar{C} + k$, (k a constant). In this case corresponding curves are parallel, i. e., one may be obtained from the other by a shift in the y^3 direction.

Furthermore, if S is an exsurface of the form

$$y^1 = x^1,\ y^2 = x^2,\ y^3 = f(x^1, x^2, m), \tag{11.16}$$

$P(a, b)$ a given base point, A, B, and C the quantities $\partial f(a, b, m)/\partial x^1$, $\partial f(a, b, m)/\partial x^2$, and $f(a, b, m)$ respectively (m variable), then the pseudo plane determined by

$$P_{y^3} = A \cdot (x^1 - a) + B \cdot (x^2 - b) + C \tag{11.17}$$

is by definition the tangent pseudo plane associated with P. Evidently, if we expand y^3 in a Taylor's series about P, we have, for each value of m,

$$y^3 = P_{y^3} + \text{higher terms.} \tag{11.18}$$

Consequently, we may say that each base curve through $(a, b,)$ determines a pseudo plane curve $C(P)$ and an exsurface curve $C(S)$ and that $C(P)$ and $C(S)$ have in common the point a, b, y^3 (a, b, m) and the tangent vector $x'^1, x'^2, dC(m)/dt + Ax'^1 + Bx'^2$.

12. Exsurfaces of the Synge type. In the case of the generalization of class one defined by equation (3.2) it should be observed that the fundamental metric relations assume a new form. For example, $(ds/dt)^2$ is the sum of a "quadratic form," a "linear form," and a scalar. For denoting

$$\sum_{i=1}^{L} \partial y^i/\partial x^{(\alpha)a} \cdot \partial y^i/\partial x^{(\beta)b} \quad \text{by} \quad g_{\alpha a . \beta b},$$

$$\sum_{i=1}^{L} \partial y^i/\partial t \cdot \partial y^i/\partial x^{(\alpha)a} \quad \text{by} \quad g_{\alpha a},$$

and

$$\sum_{i=1}^{L} \partial y^i/\partial t \cdot \partial y^i/\partial t \quad \text{by} \quad g,$$

we have

$$(ds/dt)^2 = g_{\alpha a . \beta b} x^{(\alpha+1)a} x^{(\beta+1)b} + 2g_{\alpha a} x^{(\alpha+1)a} + g. \tag{12.1}$$

Similarly, the dot product of the tangent vectors dy_1/dt, dy_2/dt of two curves is given by

$$dy_1/dt \cdot dy_2/dt = g_{\alpha a . \beta b} x_1{}^{(\alpha+1)a} x_2{}^{(\beta+1)b} + g_{\alpha a}(x_1{}^{(\alpha+1)a} + x_2{}^{(\alpha+1)a}) + g. \tag{12.2}$$

TEXAS UNIVERSITY.